

Inhoud

Artikelen

<i>Reeks 'Werving en Selectie'</i> Terug van nooit weggeweest Inleiding op de nieuwe reeks 'Werving en Selectie' <i>Eva Derous, Mandy E.G. van der Velde & Marise Ph. Born</i>	3
Instrumenten om Personeel te Selecteren in de 21ste Eeuw: Onderzoek en Praktijk <i>Filip Lievens & Britt De Soete</i>	18
<i>Reeks 'Psychologische instrumenten'</i> Werkverslaving, een begrip gemeten <i>Wilmar Schaufeli, Corine van Wijhe, Maria Peeters & Toon Taris</i>	43
Ervaren Uitdaging op het Werk Ontwikkeling van een Meetinstrument <i>Paul T. Y. Preenen, Annelies E. M. van Vianen, Irene E. de Pater & René Geerling</i>	64
Boekbespreking Prestatiemeting van wetenschappelijk onderzoek	84
Reviewers Jaargang 2010	92
Auteursregister Jaargang 2010	94
Summaries	97



ARTIKELLEN

REEKS 'WERVING EN SELECTIE'

Terug van nooit weggeweest

Inleiding op de nieuwe reeks 'Werving en Selectie'

*Eva Deros, Mandy E.G. van der Velde & Marise Ph. Born**

Nederland en Vlaanderen hebben van oudsher een goede reputatie op het gebied van onderzoek en nieuwe ontwikkelingen binnen het domein van de werving en selectie. Mede vanwege recente ontwikkelingen op de arbeidsmarkt vragen dienstverlenende bedrijven die actief zijn op het gebied van Human Resources (HR) om competente werving- en selectieconsulenten. Echter, de aandacht voor werving en selectiethema's is het laatste decennium minder sterk aanwezig in de Nederlandstalige onderzoeksliteratuur. Het aanbod van gespecialiseerde selectievakken binnen het universitair onderwijs en hoger beroepsonderwijs is mager en een inhoudelijke bespreking van werving- en selectiethema's binnen diverse HR-curricula blijkt eveneens beperkt. Deze bevindingen staan in contrast met ontwikkelingen op de arbeidsmarkt, alsook met de behoefte aan wetenschappelijke inzichten inzake nieuwe praktijken en methoden. De nieuwe reeks 'Werving en Selectie' biedt een forum aan onderzoekers en 'praktijkmensen' om tegemoet te komen aan deze nieuwe uitdagingen. In deze inleiding op de nieuwe reeks bespreken we zes trends, de kloof tussen wetenschap en praktijk, alsook enkele perspectieven om tegemoet te komen aan diverse nieuwe ontwikkelingen binnen dit domein.

1 Inleiding

Gedrag & Organisatie presenteert een nieuwe, doorlopende reeks 'werving en selectie', waarin aandacht besteed wordt aan recente ontwikkelingen op het terrein van Werving en Selectie binnen het Nederlandstalige vakgebied. Het onderscheid tussen werving en selectie is niet altijd even duidelijk. Wij verstaan onder werving alle activiteiten die betrekking hebben op het aantrekken van potentiële medewerkers. Onder selectie verstaan we alle activiteiten die betrekking hebben op het screenen van potentiële medewerkers (extern of intern). In de bijdragen

* Eva Deros is werkzaam aan de Universiteit Gent, vakgroep Personeelsmanagement, Arbeids- en Organisationspsychologie. Mandy E.G. van der Velde is werkzaam aan de Universiteit Utrecht. Marise Ph. Born is werkzaam aan de Erasmus Universiteit Rotterdam.

voor dit thema vragen wij aandacht voor de recente ontwikkelingen binnen dit vakdomein.

De behoefte aan wetenschappelijk onderbouwde en maatschappelijk verantwoorde werving- en selectiesystemen is groot. Werknemers vormen de kern van organisaties; het kiezen en selecteren van personeel blijft cruciaal voor het goed functioneren van organisaties. Een adequaat werving- en selectiebeleid bepaalt niet alleen de kwaliteit van de instroom, maar zet ook de richtlijnen voor een goede doorstroming en een goed personeelsbeleid in organisaties.

Sinds het begin van de 20ste eeuw gaat binnen de arbeids- en organisatiepsychologie veel onderzoeks aandacht uit naar het vraagstuk van werven en selecteren van medewerkers (zie Guion, 1965; Thorndike, 1949). De selectiepsychologie is immers één van de oudste en belangrijkste toepassingsgebieden van de psychologie (Hofstee, 1970; 1983) en vormt een kernonderdeel van de arbeids- en organisatiepsychologie. Het Occupational Net (O*NET; Peterson et al., 1999) geeft zelfs als belangrijkste kerntaak van de praktiserende arbeids- en organisatiepsycholoog het ontwikkelen en implementeren van selectie- en plaatsingsprocedures.

Ook Nederlandstalige onderzoekers hebben van oudsher belangrijke bijdragen geleverd aan de ontwikkeling van het werving- en selectiedomein. Initiatieven zoals de oprichting van de Commissie Testaangelegenheden Nederland in 1959 (COTAN; Evers, Lucassen, Meijer & Sijtsma, 2009) en standaardwerken zoals *De psychologische test* (Drenth, 1965) en *Testtheorie* (Drenth & Sijtsma, 2006), *Grondslagen van de Personeelselectie* (Roe, 1983), *The Blackwell Handbook of Personnel Selection* (Evers, Anderson & Voskuil, 2005), en *Naar een nieuwe generatie assessment* (Lievens & Schollaert, 2008) getuigen van een lange en gefundeerde onderzoekstraditie in Nederland en Vlaanderen.

Werving en selectie nemen de laatste jaren echter een minder prominente plaats in binnen het Nederlandstalige onderzoeksdomein. Van de 15 meest bestudeerde onderzoeksthema's die de afgelopen 10 jaar gepubliceerd zijn in *Gedrag & Organisatie* komen werving- en selectiethema's slechts op de 7de plaats (Smulders & Schouteten, 2009). Een traditioneel onderzoeksthema zoals werving en selectie eindigt dus in de middenmoot. Na een periode van bloeiende onderzoeksgroepen die met name aan de Vrije Universiteit Amsterdam onder Pieter Drenth en aan de Rijksuniversiteit Groningen onder Willem Hofstee tot in de jaren '90 floreerden, is de onderzoeks aandacht sterker komen te liggen op het thema arbeid en gezondheid, en toegepaste sociaalpsychologische thema's in organisaties, zoals leiderschap en werken in teams.

Is er de laatste jaren daardoor ook minder aandacht gekomen voor de werving- en selectieproblematiek binnen de traditionele onderwijscurricula aan Nederlandse en Vlaamse universiteiten? Een recente rondvraag bij hoogleraren en hoofd-docenten arbeids- en organisatiepsychologie aan Nederlandse en Vlaamse universiteiten leert dat werving en selectie veelal besproken wordt in het kader van algemene cursussen Personeelsbeleid, Human Resources Management (HRM) en/of inleidende cursussen Arbeids- en Organisationspsychologie. Slechts 45% van de 11 bevraagde curricula Arbeids- en Organisationspsychologie en/of Sociale en Organisationspsychologie biedt een cursus aan die integraal en exclusief gewijd is aan

werving en selectie. Slechts twee opleidingen richten daarnaast een specialisatievak personeelsselectie voor masterstudenten in. Opmerkelijk, men vermeldt tevens dat een aanzienlijk deel van de pas afgestudeerde arbeids- en organisatiepsychologen de loopbaan start als adviseur of consultant in werving en selectie. Het relatief beperkte aantal specialisatievakken werving en selectie binnen de huidige curricula aan Nederlandse/Vlaamse universiteiten staat in schril contrast met de behoefte op de arbeidsmarkt aan competente en wetenschappelijk opgeleide werving- en selectieconsulenten (Ryan & Tippins, 2004). Niet alleen arbeids- en organisatiepsychologen maar ook andere psychologen en niet-psychologen zijn actief binnen het domein van de arbeidsbemiddeling, zoals klinisch psychologen, juristen en economen. Diverse werving- en selectiekantoren bieden 'on-the-job' trainingen aan, onder andere als gevolg van de eerder beperkte focus op werving en selectie in de verschillende Human Resources (HR)-opleidingen. Opvallend is dat uit onze rondvraag blijkt dat minstens twee business schools in Nederland en Vlaanderen wél gespecialiseerde cursussen aanbieden die zich met name richten op niet-psychologen. Soms nemen ook koepelfederaties deze taak over. Zo startte de voorloper van het huidige, Belgische Federgon¹ in het midden van de jaren '90 het Young Academy Program op, een specifiek opleidingsprogramma voor werving- en selectieconsulenten. Vanaf februari 2011 lanceert Federgon opnieuw een opleidingsprogramma (m.n. het postgraduaat Recruitment, Search and Selection) in samenwerking met ervaren consultants en docenten van diverse Vlaamse universiteiten. Deze opleiding combineert een praktijkgerichte opleiding met een stevige theoretische onderbouwing om gevolg te geven aan de groeiende vraag naar een verdere professionalisering van de HR-dienstverlening en ontwikkelingen op de arbeidsmarkt.

Het is opvallend dat in de Verenigde Staten de selectiepsychologie vanouds en nog steeds één van de belangrijke thema's binnen de arbeids- en organisatiepsychologie vormt. Het jaarlijkse Amerikaanse congres van SIOP, de Society for Industrial and Organizational Psychology, besteedt traditioneel veel aandacht aan werving en selectie en in bredere zin ook aan personeelspsychologische thema's. Deze aandacht in de Verenigde Staten past in de van oudsher meritocratische samenleving waar men erop uit is het grootste talent te vinden en te selecteren. Bovendien doen in dat land veranderingen zoals technologische ontwikkelingen, veranderingen in de wervings- en selectiepopulatie en ook test- en meettheoretische ontwikkelingen, zich vaak in een sneller tempo voor dan in Nederland en Vlaanderen. In het huidige tijdsgewricht in Nederland en Vlaanderen, dat duidelijk meritocratischer is dan enkele decennia geleden, heeft dit domein van de psychologie veel te bieden. Het is zeker de moeite waard meer aandacht aan de thematiek van werving en selectie te gaan geven. In wat volgt, bespreken we enkele actuele vragen voor onderzoek en praktijk. Daarnaast bekijken we enkele perspectieven op werving en selectie om tegemoet te komen aan nieuwe uitdagingen.

2 Actuele vragen voor onderzoek en praktijk

Sociaaleconomische en maatschappelijke veranderingen worden snel opgepakt door de arbeidsmarkt. Ook de werving- en selectiepraktijk is onderhevig aan deze veranderingen. Zo functioneert de uitzendsector als graadmeter voor economische impasse of hausse en ontsnappen wervers en selecteurs niet aan het hedendaagse reveil van het ethisch ondernemen. De laatste decennia worden gekenmerkt door een toenemende diversiteit (ontgroening, vergrijzing, verkleuring en vervrouwelijking), technologische veranderingen en snelheid en competitiviteit van ondernemen. Nieuwe thema's, zoals ethisch verantwoord ondernemerschap, komen terug in beroepscode, zoals de sollicitatiecode van de Nederlandse Vereniging voor Personeelsmanagement en Organisatieontwikkeling (NVP, 2006) en de professionele richtlijnen voor het omgaan met sollicitanten (bijv. Bochhah, Kort & Seddik, 2005). Een goed personeelsbeleid probeert adequaat in te spelen op nieuwe uitdagingen. In wat volgt, schetsen we zes actuele trends en daaruit voortvloeiende vragen en uitdagingen voor het onderzoek en de praktijk van het werven en selecteren. Deze zes trends zijn niet uitputtend, maar bij wijze van voorbeeld bedoeld.

2.1. Voorspellen van verandering

Het voorspellen van verandering binnen selectie lijkt een 'contradictio in terminis'. Het traditionele selectiemodel is een predictief validiteitsmodel dat steunt op de klassieke functieanalyse, waarbij predictoren en criteria als relatief stabiel worden opgevat. In een sterk veranderende arbeidsmarkt komt het traditionele selectiemodel onder druk te staan. Criteria worden dynamischer. Functies worden minder vastgelegd en steeds meer ingevuld door het individu (bijv. 'job crafting', zie Wresniewski & Dutton, 2001). Deze veranderingen gelden niet voor alle type banen en sectoren, maar vinden we vooral terug in bedrijven die gekenmerkt worden door een kenniseconomie, waarbij manuele arbeid uitbesteed wordt. Door de kenniseconomie verandert ook het type psychologisch contract tussen de organisatie en (potentiële) medewerker. Onder psychologisch contract verstaan we het geheel van wederzijdse, veelal ongeschreven verwachtingen die tussen elke medewerker, manager en significante andere in de organisatie spelen. Studies naar het psychologisch contract beginnen doorgaans met het moment waarop de nieuwe medewerker een mondelinge of schriftelijke verbintenis aangaat en zijn/haar intrede doet in de organisatie. Maar vóór het eigenlijke contract ondertekend wordt, tasten sollicitant en organisatie al hun grenzen af. Wat kan de sollicitant de organisatie bieden? Wat kan de organisatie de sollicitant bieden? Klikt het tussen de sollicitant en de organisatie? Vanaf het doornemen van advertenties en het eerste gesprek tot en met de formele contractonderhandelingsfase worden verwachtingen gecreëerd en uitgewisseld. Uit onderzoek blijkt dat verwachtingen over de inzetbaarheid van individuele medewerkers belangrijker worden dan levenslange loyaliteit en werken bij dezelfde werkgever. Het gevolg hiervan is dat men een nieuw soort competenties hanteert die gekenmerkt worden door veranderingsbereidheid, flexibiliteit, creativiteit, leervermogen, en informatieverzameling en -verwerking. De selectielat wordt hierdoor hoog gelegd: niet

iedereen voelt zich geroepen of is in staat om levenslang inzetbaar te blijven en zich aan veranderende omstandigheden aan te passen. Daarnaast blijft er conceptuele verwarring bestaan over de construct- en criteriumgerelateerde validiteit van competenties, zoals innovativiteit en creativiteit (bijv. Sternberg & Lubart, 1999). Kortom, als functie-inhoud (en dus criteria) snel veranderen, kan er dan nog wel adequaat voorspeld worden? Het blijft een uitdaging om predictie af te stemmen op verandering. Nieuwe competenties en benaderingen dringen zich op, zoals het inkorten van werving- en selectiecycli en het permanent assessment van personeel (bijv. via development centers en training centers; zie Hoekstra & Van Sluijs, 2003).

2.2 Diversiteit

Onder invloed van globalisering, het vrijmaken van markten, demografische ontwikkelingen en re-integratie van minderheidsgroepen, wordt de beroepsactieve bevolking in West-Europa diverser. Er is sprake van 'ontgroening' (d.w.z. een afname van het aantal jongeren in de beroepsactieve bevolking), 'vergrijzing' (d.w.z. een toename van het aantal ouderen) en 'verkleuring' (d.w.z. een toename van diverse etnische groepen op de arbeidsmarkt). Sollicitanten uit de zogenaamde 'minderheidsgroepen' worden actief op de arbeidsmarkt en brengen nieuwe uitdagingen mee. Selectie-instrumenten zijn van oudsher ontwikkeld voor, en gevalideerd in, traditionele (autochtone) bevolkingsgroepen waardoor bestaande tests onderhevig kunnen zijn aan zogenaamde 'adverse impact'. Adverse impact is een technische term die verwijst naar een differentiële behandeling van leden uit minderheidsgroepen ten aanzien van leden uit meerderheidsgroepen (Outtz, 2009). In het bijzonder is er sprake van adverse impact wanneer de selectieratio van een meerderheidsgroep (bijvoorbeeld autochtone sollicitanten) substantieel groter is dan de selectieratio van een minderheidsgroep (bijvoorbeeld allochtone sollicitanten). Adverse impact kan te wijten zijn aan diverse factoren waaronder ware verschillen tussen minderheden- en meerderheden-groepen in functierelevante competenties, maar ook aan testbias (zowel op test- als itemniveau). Daarnaast kan sprake zijn van onterechte discriminatie te wijten aan beoordelingsfouten en stereotypering (bijv. Deros, Nguyen & Ryan, 2009). In de Verenigde Staten beschouwt de rechter adverse impact als een 'presumption of discrimination', tenzij de werkgever quota hanteert, functierelevantie en/of het gebrek van alternatieve tests aan kan tonen. Adverse impact moet dus vermeden en/of weerlegd worden door de werkgever op straffe van juridische sancties. In Europa is onterechte discriminatie bij aanwerving eveneens verboden. De vraag die zich opdringt, is of de Nederlandse/Vlaamse wetgever een strengere stramen en een duidelijker beslissingsboom kan hanteren om onterechte discriminatie op te sporen en tegen te gaan zoals in de Verenigde Staten het geval is. Bijvoorbeeld, in Vlaanderen is veel discussie over het belang en de rechtsgeldigheid van situatie-tests om onterechte discriminatie bij aanwerving op te sporen. In Nederland stelden Bochhah et al. (2005) een aantal richtlijnen op voor het diagnosticeren van etnische minderheden. Bij veel tests zijn nog te veel ethnocentrische en cultuurspecifieke elementen te vinden, hoewel er ook inspanningen worden verricht voor het ontwikkelen van multicultureel toepasbare tests en vragenlijsten. Een belang-

rijke vraag die onderzoekers en practici bezighoudt, is de spanning tussen maximalisatie van gelijke kansen (diversiteit) en maximalisatie van doelmatigheid (validiteit), waarbij er een 'trade-off' is tussen beide principes. Met andere woorden: de tests die het meest predictief valide zijn (maximalisatie van doelmatigheid) leiden veelal tot de meeste adverse impact (minimale gelijke kansen). Dit zogenaamde diversiteit-validiteitsdilemma (Ployhart & Holtz, 2008) krijgt momenteel veel aandacht (Born, 2010).

2.3 Contextualisatie

Een andere, recente thematiek heeft te maken met het contextualiseren van werving- en selectiepraktijken. Contextualisatie verwijst naar de ontwikkeling en het gebruik van functie-relevante en contextspecifieke tools (Schmit, Ryan, Stierwalt & Powell, 1995). Predictoren (bijvoorbeeld testitems die cijfermatig inzicht meten) worden aangepast aan de te voorspellen criteria (bijv. redeneren met economische gegevens). Er wordt veel voordeel verwacht van het contextualiseren van tests (Lievens, De Corte & Schollaert, 2008), zoals een meer positieve testperceptie en testmotivatie bij kandidaten, alsook een verhoogde voorspellende validiteit en verminderde adverse impact (bijv. door het gebruik van gedragsproeven en work sample tests). Zo blijft een belangrijke uitdaging hoe geselecteerd kan worden voor diversiteit en in hoeverre werving- en selectiepraktijken zich dienen aan te passen aan specifieke doelgroepen inzake demografische kenmerken (zoals sekse, leeftijd, etniciteit) en/of met betrekking tot specifieke taken en opdrachten (zoals expatriation; Mol, Born, Willemsen, Van der Molen & Deros, 2009). Kortom, hoe ver kunnen we gaan met het contextualiseren van werving- en selectietools zodat de resultaten (bijv. validiteit) in verhouding staan tot de opbrengsten (bijv. diversiteit) en de zware ontwikkelingsinspanningen?

2.4 Technologische innovatie

Steeds vaker worden multimediatechnieken ingeschakeld om kandidaten te werven en te selecteren. Sinds het begin van de jaren '60 worden traditionele 'paper-and-pencil' tests vervangen door geautomatiseerde tests en procedures (Bartram & Bayless, 1984) die onder andere gebruik maken van moderne multimedia-technieken. Multimedia tests worden gekenmerkt door video- en audiomateriaal om sollicitanten te screenen. Voorbeelden van multimedia tests zijn multimedia situational judgment tests (SJT's, bijv. Olson-Buchanan & Drasgow, 2006; en webcamtests, bijv. Ostrom, Born, Serlie & Van der Molen, 2010), interviews via videoconferentie (bijv. Chapman, Uggerslev & Webster, 2003), en tests die gebruik maken van virtuele realiteit (bijv. Aguinas, Henle & Beaty Jr., 2001). Multimedia tests laten een actieve, dynamische uitwisseling van informatie tussen organisaties en kandidaten toe (Chapman & Webster, 2003). Door middel van multimediatoepassingen kunnen vaardigheden en persoonlijkheidskenmerken van kandidaten gemeten worden die beperkt of moeilijker meetbaar zijn door middel van traditionele tests. Voorbeelden zijn interpersoonlijke vaardigheden, zoals integriteit, conflicthantering en leidinggeven. Omgekeerd, kunnen kandidaten via websites en filmpjes sneller een concreet beeld krijgen van de baaninhoud ('realistic job preview') en de cultuur van de organisatie waar men wil solliciteren.

Niet alleen organisaties, maar ook sollicitanten kunnen zich presenteren door middel van nieuwe technieken, zoals video-cv's.

HR-professionals koesteren hoge verwachtingen van de multimediatoepassingen, waaronder een hogere realiteitswaarde ('fidelity'), een ingekorte werving- en selectiecyclus, minder adverse impact, een lager kostenplaatje, een positief bedrijfsimago en positieve reacties van sollicitanten op de gehanteerde werving- en selectiepraktijken (Chapman & Webster, 2003). Gegeven deze hoge verwachtingen is het niet verwonderlijk dat multimediatechnieken in snel tempo geïmplementeerd worden in organisaties. Of de multimediatechnieken aan deze verwachtingen kunnen voldoen, is minder duidelijk. Het onderzoek naar effecten van multimediatechnieken op de organisatie van werving- en selectiepraktijken, met inbegrip van online sociale netwerken en mediaplatformen (zoals LinkedIn en Twitter) staat nog in de kinderschoenen. Lievens en Harris (2003) pleitten voor meer onderzoek naar psychometrische kenmerken, equivalentie en reacties van sollicitanten op multimedia tools. Een aantal Nederlandse/Vlaamse onderzoekers is recent gestart met het beantwoorden van deze 'call to action' (Hiemstra, Derous, Born & Serlie, 2011; Oostrom, 2010).

2.5 *Ethisch reveil*

Jérôme Kerviel van de Franse bank Société Générale werd in januari 2008 door Société Générale verantwoordelijk geacht voor het recordverlies van € 4,9 miljard en in 2010 veroordeeld tot een celstraf van vijf jaar. Kerviel bekende dat hij documenten vervalste en het financieel controlesysteem omzeilde. Meer algemeen is contraproductief en onethisch organisatiegedrag de laatste jaren door de financiële crisis sterker in de maatschappelijke belangstelling komen te staan en integriteit maar ook risicovolle persoonlijkheidseigenschappen krijgen meer aandacht bij werving en selectie. Mensen met een externe 'locus of control' zien anderen of het toeval als verantwoordelijk voor situaties om zich heen (Rotter, 1954). Bandura (1999) bespreekt rationalisaties die verhinderen dat de zelfregulerende mechanismen tot moreel handelen leiden maar wel in zogenaamd moreel disengagement resulteren. In lijn met de belangstelling voor ethiek en moraliteit, doen Van Oudenhoven, Blank, Leemhuis, Pomp en Sluis (2008) de suggestie voor het ontwikkelen van een deugden-persoonlijkheidstest. Recent toonden Ashton en Lee (2005) aan dat er niet vijf (zoals algemeen aangenomen door de Big Five persoonlijkheidsmodellen), maar zes relatief onafhankelijke persoonlijkheidsdimensies zijn, zoals weergegeven in het zogenaamde HEXACO-model. HEXACO is een acroniem dat verwijst naar Honesty-Humility (Integriteit), Emotionality (Emotionaliteit), eXtraversion (Extraversie), Agreeableness (Verdraagzaamheid), Conscientiousness (Consciëntieusheid) en Openness to experience (Openheid voor ervaringen). De zesde dimensie, integriteit, verwijst naar iemands geneigdheid om anderen niet te manipuleren voor persoonlijk gewin en regels te overtreden. Integriteit verwijst verder naar iemands relatieve desinteresse in luxe, status en privileges (De Vries, Ashton & Lee, 2009). Onderzoek van De Vries et al. (2009) bevestigt de 6-factorenstructuur van persoonlijkheid en het bestaan van een 'integriteitsdimensie' binnen de Nederlandse context.

Heeft het echter zin om locus of control, moreel redeneren en/of integriteit als voorspeller bij selectieprocedures te gebruiken? In hoeverre kunnen we contra-productief werkgedrag bij intrede reeds voorspellen? Een belangrijke, hieraan gerelateerde vraag is hoe integriteit het best gemeten kan worden (zie Lievens & De Soete, dit nummer). Daarnaast speelt naast een zekere 'voorbesteding' (zoals vastgelegd in iemands persoonlijkheid en/of andere persoonskenmerken) wellicht ook de werkcontext een belangrijke rol bij het tot uiting komen van (im)moreel, (on)eerlijk en (contra)productief werkgedrag (bijv. Cem Ersoy, Born, Deros & Van der Molen, 2010). Onderzoek naar de construct- en criteriumgerelateerde validiteit van het meten van risicovolle persoonlijkheidseigenschappen voor het voorspellen van werkgedrag verdient dan ook meer aandacht.

2.6 Ontwikkelingen in de psychometrie

Binnen de psychometrie valt de laatste jaren een tendens waar te nemen waarbij de item respons theorie (IRT) nadrukkelijker wordt gebruikt om de kwaliteit van niet-cognitieve meetinstrumenten na te gaan. De IRT is van oudsher vooral toegepast op het cognitieve domein. Het gebruik van testcores voor belangrijke personeelsbeslissingen betekent dat het noodzakelijk is dat ook niet-cognitieve tests zoals persoonlijkheidstests van goede kwaliteit zijn. Aan de hand van IRT-bevindingen blijkt dat persoonlijkheidsschalen vaak psychometrisch zwak zijn en onvoldoende precies meten, of erg smalle metingen zijn van het bedoelde begrip. Ook kan de IRT licht werpen op inconsistente antwoordpatronen en afwijkend antwoordgedrag op vragenlijsten. Er valt nog veel te winnen met de inzichten vanuit de psychometrie, zoals het gebruik van IRT voor het verbeteren van assessments in het niet-cognitieve domein (bijv. Egberink, 2010; Egberink, Meijer, Veldkamp, Schakel & Smid, 2010).

3 Perspectieven op werving en selectie

Zoals aangegeven, zijn werving- en selectieprocedures onderhevig aan sociaal-economische en maatschappelijke bewegingen. Zowel Amerikaans onderzoek (bijv. Rynes, Colbert & Brown, 2002) als Europees onderzoek (bijv. König, Klehe, Berchthold & Kleinmann, 2010) wijzen echter op een groeiende kloof tussen het wetenschappelijke onderzoek en de praktijk, of de zogenaamde 'scientist-practitioner gap'. Zo worden, bijvoorbeeld, de betrouwbaarheid en voorspellende waarde van het ongestructureerde interview steevast in twijfel getrokken door academici (McDaniel, Whetzel, Schmidt & Maurer, 1994; Schmidt & Hunter, 1994). Toch blijft het ongestructureerde interview wereldwijd een van de meest gehanteerde selectie-instrumenten. Volgt de praktijk de academici niet of volgen de academici de praktijk niet? Hoe kunnen we de kloof tussen wetenschappelijk onderzoek en praktijk begrijpen? Verschillende factoren spelen een rol. Zo wordt een groot aantal onderzoeksbevindingen gepubliceerd in voor de praktijk weinig toegankelijke tijdschriften. Sommige tijdschriften zijn bovendien erg prijzig. Ook de voertaal en het typische academische jargon kunnen afschrikken. Gedrag & Organisatie wil hieraan tegemoet komen, onder meer door toevoeging van een

praktijkbox, waarin onderzoeksbevindingen vertaald worden naar concrete aanbevelingen voor de praktijk. Daarnaast zijn sommige onderzoeksbevindingen moeilijk generaliseerbaar vanwege het beperkt aantal contextfactoren dat bestudeerd kan worden (zoals budgettaire beperkingen en tijdsbeperkingen). Bovendien is de praktijk veelal te complex om in één of meerdere studies gevat te worden. Anderzijds werken veel praktijkmensen onder enorme druk (tijdsdruk, klant-eisen en financiële streefdoelen), waardoor men nieuwe, wetenschappelijke inzichten en methoden niet altijd kan toepassen of mag uittesten. Of ligt laatstgenoemde aan het feit dat practici vasthouden aan de eigen intuïtie en subjectiviteit wanneer ze kandidaten beoordelen? (Highhouse, 2008). Bijvoorbeeld, vele interviewers zijn ervan overtuigd dat ervaring en intuïtie het verschil maken tussen goede en minder goede beoordelingen van sollicitanten. Echter, zelfs bij een perfect betrouwbaar en valide criterium voorspellen interviews doorgaans niet meer dan 10% van de variantie in werkgedrag. De overtuiging dat klinische predictie gebaseerd op ervaringsdeskundigheid en gevoel leidt tot betere oordelen is zeer populair, maar wordt thans niet ondersteund door onderzoeksbevindingen.

De uitdaging voor academici blijft om adequate antwoorden te bieden op nieuwe ontwikkelingen die tevens relevant zijn voor de praktijk. Gedurende decennia hebben academici de personeelsselectie hoofdzakelijk vanuit een klassiek, psychometrisch perspectief bestudeerd. Dit perspectief vertrekt vanuit het individuele verschilparadigma (intelligentie; persoonlijkheid), legt de nadruk op stabiliteit (wat betreft functies, functieomgeving en functie-uitvoerders) en veronderstelt een eenrichtingsverkeer tussen werkgever/selecteur enerzijds en sollicitanten anderzijds. Zowel de economische crisis van de jaren '30 als de beide wereldoorlogen zijn belangrijke impulsen geweest voor de verdere ontwikkeling van het klassieke selectiemodel. De opkomst van de decisietheorie in de jaren '50-'60 en de behoefte aan kostenbesparing vormden de voedingsbodemp voor het utiliteitsdenken binnen de personeelsselectie. Het klassieke perspectief op werving en selectie is bovenal een functioneel perspectief: het is de bedoeling om de juiste persoon op de juiste plaats te krijgen op een zo efficiënt mogelijke manier. Utiliteitsstudies tonen aan dat werving- en selectieprocedures onlosmakelijk verbonden zijn met het succes van individuele medewerkers c.q. organisaties (bijv. Judiesch, 2001). Anderzijds houdt het klassieke selectiemodel weinig tot geen rekening met de dynamische en complexe aard van gedrag in organisaties. Het gedrag van een persoon is niet alleen een functie van zijn/haar capaciteiten (bijv. intelligentie, persoonlijkheid), maar ook van de complexe interactie tussen de hele persoon en diens omgeving. Factoren zoals testmotivatie, testangst, testvoorkeuren, vermoeidheid, en gepercipieerde rechtvaardigheid van selectieprocedures zijn inherent verbonden met het testgedrag en bepalen de complexiteit van assessmentprocedures. Echter, binnen de klassieke selectiepsychologie worden laatstgenoemde factoren hoogstens als bron van niet-systematische foutenvariantie beschouwd.

Onderzoek naar werving- en selectiepraktijken vergt niet alleen de nodige inventiviteit, maar ook moed om tegemoet te komen aan het 'scientist-practitioner' perspectief. Vanuit wetenschappelijk oogpunt vraagt het onderzoek naar diverse werving- en selectiethema's veelal andere, meer complexe designs en meetmetho-

des dan, bijvoorbeeld, het klassieke cross-sectionele vragenlijstonderzoek. Verder dient men zorgzaam om te springen met de relatief schaarse onderzoeksmogelijkheden en -settings. Onderzoek onder studentenpopulaties is bijvoorbeeld vrij makkelijk uitvoerbaar maar wordt niet snel aanvaard binnen het toegepaste domein van de werving en selectie. Selectie blijft een gevoelig thema. In vergelijking met traditioneel onderzochte werknemerspopulaties binnen de arbeids- en organisatiepsychologie zijn sollicitantenpopulaties en werkgevers moeilijker te overtuigen om mee te werken aan onderzoek. Bovendien kunnen de 'high-stake' selectieomstandigheden (statistische) beperkingen opleggen, en kunnen criteriumcontaminatie en uitval onder sollicitanten (bijv. bij predictieve validiteitsstudies) de validiteit van de onderzoeksbevindingen ondergraven.

Veranderingen op de arbeidsmarkt en de kloof tussen wetenschap en praktijk leidden de laatste twee decennia tot het formuleren van een alternatief model: het sociaal procesmodel van werving en selectie (Deros & De Witte, 2001; Herriot, 1989; Iles, 1999). Dit model vertrekt vanuit andere basisveronderstellingen dan het klassieke model, zoals flux en transformatie (qua persoonlijke functioneren, functies en functieomgevingen), een sociaal-motivatieve kijk op (test)gedrag en andere of veranderende machtsverhoudingen omwille van schaarste op de arbeidsmarkt. Binnen deze sociale procesbenadering is er bovenal aandacht voor een tweerichtingsverkeer: niet alleen de organisatie maar ook de kandidaat verzamelt informatie om tot een beslissing te komen. Werving- en selectiebeslissingen zijn dus het resultaat van beslissingen die zowel organisaties als kandidaten nemen. Aangezien het standpunt van sollicitanten binnen het traditionele model onderbelicht bleef, hebben onderzoekers zich de laatste decennia toegelegd op percepties en reacties van kandidaten op sollicitatieprocedures (bijv. hoe rechtvaardig procedures zijn) en – bij uitbreiding – diverse organisatievariabelen (bijv. organisatieattractiviteit), alsook het zelfbeeld en de zelfwaardering van kandidaten (Hausknecht, Day & Thomas, 2004). Organisaties zijn bezorgd om de percepties van sollicitanten (König et al., 2010), maar de relatie tussen testpercepties en gedragsmatige output op individueel niveau (zoals testprestaties, uitvalgedrag) en organisatieniveau (zoals Organizational Citizenship Behavior of OCB) blijft onderbelicht. Tot nu toe vallen de onderzoeksresultaten wat mager uit (Sackett & Lievens, 2008). De uitdaging bestaat erin om deze relaties verder te onderzoeken (zie bijv. Schreurs, Deros, Van Hooft, Proost & De Witte, 2009). Indien dat niet gedaan wordt dan blijft het meten van sollicitantvoorkeuren inzake werving- en selectieprocedures een maat voor niets en – paradoxaal genoeg – bijdragen aan de kloof tussen wetenschap en praktijk.

Het klassieke model en het sociale procesmodel lijken diametraal tegenover elkaar te staan. Tot op zekere hoogte is dat zo, aangezien beide modellen vanuit andere basisassumpties vertrekken. Maar beide modellen illustreren vooral dat werving en selectie complex is: 'good practice' zal zich ergens in het midden bevinden en afhankelijk zijn van de concrete vraag en/of context. Bijvoorbeeld, in tijden van arbeidsmarktschaarste worden werving en arbeidsmarktcommunicatie belangrijker dan doorgedreven predictie. Bij overaanbod op de arbeidsmarkt, daarentegen, zal predictie meer de doorslag geven. We pleiten daarom voor onderzoek dat beide perspectieven tracht te integreren en af te stemmen. Recent onderzoek

naar het diversiteit-validiteitsdilemma is een concrete illustratie van de spanning die tussen beide perspectieven bestaat en van de manieren om hieraan tegemoet te komen door beide perspectieven op elkaar af te stemmen (bijv. Lievens & De Soete, dit nummer; Outtz, 2009).

4 Besluit

In deze bijdrage besteedden we aandacht aan de vraag hoe het staat met het werving- en selectieonderzoek in Nederland en Vlaanderen. Uit een recente overzichtstudie (Smulders & Schouteten, 2009) blijkt dat het thema de laatste jaren minder aandacht heeft gekregen, althans wat betreft het aantal publicaties in wetenschappelijke tijdschriften, zoals *Gedrag & Organisatie*. Deze bevinding staat haaks op de van oudsher stevige reputatie die Nederlandse en Vlaamse onderzoekers hebben opgebouwd in dit domein, recente ontwikkelingen op de arbeidsmarkt en de behoefte aan wetenschappelijk opgeleide werving- en selectieconsulenten. Minstens zes actuele trends tonen aan dat onderzoekers en praktijkmensen voor nieuwe, grote uitdagingen staan. Verschillende perspectieven op werving en selectie, zoals het traditionele en sociale procesmodel, kunnen elkaar versterken om tegemoet te komen aan deze uitdagingen om ook zo de kloof tussen wetenschap en praktijk te verkleinen. In deze nieuwe reeks 'Werving en Selectie' van *Gedrag & Organisatie* bieden we een forum aan alle onderzoekers en praktijkmensen binnen Nederland en Vlaanderen die bovenstaande uitdagingen aangaan, vanuit welk perspectief dan ook. 'Werving en selectie zijn terug, van nooit weggeweest.'

Praktijkbox

Praktijkmensen en onderzoekers moeten meer samenwerken om de kloof tussen praktijk en theorie kleiner te maken en adequate antwoorden te bieden op actuele vragen en nieuwe uitdagingen. Binnen het domein van werving en selectie anno 2011 worden onder meer de volgende trends en uitdagingen onderscheiden:

- Assessments zullen steeds meer ingaan op het voorspellen van veranderingsbereidheid, leervermogen en informatieverwerking (Trend: Voorspellen van verandering).
- Naast het bieden van gelijke kansen aan diverse doelgroepen moet ook de doelmatigheid van werving- en selectieprocedures gegarandeerd worden (Trend: Diversiteit).
- Specifieke tests kunnen ontwikkeld/ingezet worden die aangepast zijn aan een specifiek type functies, doelgroepen en/of werksituaties zonder dat dit ten koste gaat van efficiëntie en uiteindelijke opbrengsten (Trend: Contextualiseren van tests).
- De test-technische kwaliteiten en aanvaardbaarheid van (nieuwe) tests en technieken die ingezet worden, zoals videotesten en webcamtesten,

moeten kritisch onderzocht worden. De moderne testtheorie kan beter benut worden om de psychometrische kwaliteit van niet-cognitieve tests, zoals persoonlijkheidstests, te verbeteren (Trends: Technologische innovatie en toepassen van de moderne testtheorie bij het meten van persoonlijkheid).

- Nieuwe persoonskenmerken, zoals moraliteit, integriteit en risicovolle persoonlijkheid, kunnen eveneens onderzocht en opgenomen worden in de traditionele testbatterijen indien ze contraproductief werkgedrag voorspellen (Trend: Ethisch werven en selecteren).

Samenvatting

Recruitment & Selection: Back from never been gone

E. Derous, M. E. G. van der Velde & M. Ph. Born (2011). *Gedrag & Organisatie*, volume 24, January 2011, nr. 1, pp. xx-xx.

The Netherlands and Flanders are well known for their long and profound tradition in research and development on recruitment and selection practices. During the past decade, however, considerably less attention has been paid towards recruitment and selection topics, both within the Dutch/Flemish academic literature as well as at Dutch/Flemish universities (e.g., in terms of the number of specialized courses). This remarkable observation runs counter to the call for scientifically trained recruiting/selection experts as instigated by several new developments within the professional field of recruitment and assessment. This article presents six developments/trends in recruitment and selection. We further illustrate the scientific-practitioner gap as well as several perspectives on recruitment and selection to address some of the recent developments.

Noot

- 1 Federgon is de Belgische beroepsfederatie van HR-dienstverleners en vertegenwoordigt onder meer de private arbeidsbemiddelaars, waaronder werving-, search- en selectiebedrijven, outplacementbedrijven, uitzendbedrijven, en interim-managementbedrijven.

Literatuur

- Aguinas, H., Henle, C.A. & Beaty Jr. (2001). Virtual reality technology: A new tool for personnel selection. *International Journal of Selection and Assessment*, 9, 70-83.
- Ashton, M.C. & Lee, K. (2005). Honesty-humility, the Big Five, and the Five-Factor Model. *Journal of Personality*, 73, 1321-1353.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3, 193-209.
- Bartram, D. & Bayless, R. (1984). Automated testing: Past, present, and future. *Journal of Occupational Psychology*, 57, 221-237.

- Bochhah, N., Kort, W. & Seddik, H. (2005). *Richtlijnen gebruik diagnostische instrumenten bij etnische minderheden*. Rotterdam/Amsterdam, Nederland: LBR/NIP.
- Born, M.Ph. (2010). Het beoordelen van mensen voor werkgerelateerde doelen: Etniciteit, moraliteit, en subjectiviteit. *Gedrag & Organisatie*, 23, 232-256.
- Cem Ersoy, N., Born, M.Ph., Derous, E. & Van der Molen, H.T. (2010). Antecedents of organizational citizenship behavior among blue- and white-collar workers in Turkey. *International Journal of Intercultural Relations*. Advance online publication.
- Chapman, D.S., Uggerslev, K.L. & Webster, J. (2003). Applicant reactions to face-to-face and technology-mediated interviews: A field investigation. *Journal of Applied Psychology*, 88, 944-953.
- Chapman, D.S. & Webster, J. (2003). The use of technologies in the recruiting, screening, and selection processes for job candidates. *International Journal of Selection and Assessment*, 11, 113-120.
- 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, 319-342.
- Derous, E., Nguyen, H.H. & Ryan, A.M. (2009). Hiring discrimination against Arab minorities: Interactions between prejudice and job characteristics. *Human Performance*, 22, 297-320.
- De Vries, R.E., Ashton, M.C. & Lee, K. (2009). De zes belangrijkste persoonlijkheidsdimensies en de HEXACO-persoonlijkheidsvragenlijst. *Gedrag & Organisatie*, 22, 232-274.
- Drenth, P.J.D. (1965). *De psychologische test*. Arnhem, Nederland: Van Loghum Slaterus.
- Drenth, P.J.D. & Sijsma, K. (2006). *Testtheorie. Inleiding in de theorie van de psychologische test en zijn toepassingen (4de herziene druk)*. Houten, Nederland: Bohn Stafleu Van Loghum.
- Egberink, I.J.L. (2010). *Applications of item response theory to non-cognitive data* (unpublished doctoral dissertation). Universiteit Groningen, The Netherlands.
- Egberink, I.J.L., Meijer, R.R., Veldkamp, B.P., Schakel, L. & Smid, N.G. (2010). Detection of aberrant item score patterns in computerized adaptive testing: An empirical example using the CUSUM. *Personality and Individual Differences*, 48, 921-925.
- Evers, A., Anderson, N. & Voskuil, O.F. (Eds.) (2005). *The Blackwell Handbook of Personnel Selection*. Oxford, England: Blackwell.
- Evers, A., Lucassen, W., Meijer, R.R. & Sijsma, K. (2009). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: Nederlands Instituut van Psychologen/Commissie Testaangelegenheden Nederland.
- Guion, R.M. (1965). *Personnel testing*. New York: McGraw-Hill.
- 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.
- 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.
- Hiemstra, A., Derous, E., Born, M.Ph. & Serlie, A. (2011). *Ethnically diverse fairness perceptions of video resumes*. Manuscript submitted for publication.
- Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology*, 1, 333-342.
- Hoekstra, H.A. & Van Sluijs, E. (2003). *Management van competenties*. Het realiseren van HRM. Assen, Nederland: Van Gorcum.
- Hofstee, W.K.B. (1970). *Selectie van personen*. Inaugurele rede. Assen, Nederland: Van Gorcum.
- Hofstee, W.K.B. (1983). *Selectie*. Utrecht/Antwerpen: Het Spectrum.

- Iles, P.A. (1999). *Managing staff selection and assessment*. Buckingham, UK: Open University Press.
- Judiesch, M.K. (2001). Using estimates of the output productivity ratio (SDp) to improve the accuracy and managerial acceptance of utility analysis estimates. *Journal of Business and Psychology*, 14, 529-551.
- König, C.J., Klehe, U.-C., Berchthold, M. & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection methods. *International Journal of Selection and Assessment*, 18, 17-27.
- 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. & De Soete, B. (2011). Instrumenten om personeel te selecteren in de 21ste eeuw. Onderzoek en praktijk. *Gedrag & Organisatie*, 24, 18-41.
- Lievens, F. & Harris, M.M. (2003). Research on internet Recruiting and testing: Current status and future directions. In C.L. Cooper & I.T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology*, vol. 16. (pp. 131-165). Chichester, UK: Wiley.
- Lievens, F. & Schollaert, E. (2008). *Naar een nieuwe generatie assessment*. Barneveld, Nederland: Uitgeverij Nelissen.
- McDaniel, M.A., Whetzel, D.L., Schmidt, F.L. & Maurer, S.D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79, 599-616.
- Mol, S.T., Born, M.Ph., Willemsen, M.E., Van der Molen, H.T. & Derous, E. (2009). When selection ratios are high: Predicting the expatriation willingness of prospective domestic entry-level job applicants. *Human Performance*, 22, 1-22.
- NVP (2006). *Sollicitatiecode*. Gedownload op 12 november 2010 van <http://www.nvp-plaza.nl/site/nl/kennis.phtml?p=sollicitatiecode>
- Olson-Buchanan, J.B. & Drasgow, F. (2006). Multimedia situational judgment test: The medium creates the message. In J.A. Weekley & R.E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 253-278). Mahwah, NJ: Lawrence Erlbaum.
- Oostrom, J.K. (2010). *New technology in personnel selection. The validity and acceptability of multimedia tests* (unpublished doctoral dissertation). Erasmus University Rotterdam, Rotterdam, The Netherlands.
- Oostrom, J.K., Born, M.Ph., 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.
- Outtz, J.L. (Ed.) (2009). *Adverse impact: Implications for organizational staffing and high stakes selection*. New York: Routledge.
- Peterson, N.G., Borman, W.C., Hanson, M.A. & Kubisiak, U.C. (1999). Summary of results, implications for O*NET applications, and future directions. In N.G. Peterson, M.D. Mumford, W.C. Borman, P.R. Jeanneret & E.A. Fleishman (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 175-195). Washington, DC: American Psychological Association.
- Ployhart, R.E. & Holtz, B.C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61, 153-172.
- Roe, R.A. (1983). *Grondslagen der personeelsselectie*. Assen, Nederland: Van Gorcum.
- Rotter, J.B. (1954). *Social learning and clinical psychology*. New York: Prentice-Hall.

- Ryan, A.M. & Tippins, N.T. (2004). Attracting and selecting: What psychological research tells us. *Human Resource Management*, 43, 305-318.
- Rynes, S.L., Colbert, A.E. & Brown, K.G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management*, 41, 149-174.
- Sackett, P.R. & Lievens, F. (2008). Personnel Selection. *Annual Review of Psychology*, 59, 419-450.
- Schmidt, F.L. & Hunter, J.E. (1994). 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.
- Schreurs, B., Derous, E., Van Hooft, E.A.J., Proost, K. & De Witte, K. (2009). Predicting applicants' job pursuit behavior from their selection expectations: The mediating role of the theory of planned behavior. *Journal of Organizational Behavior*, 30, 761-783.
- Smulders, P. & Schouteten, R. (2009). Balans van 25 jaar arbeids- en organisatiepsychologie: Nieuwe thema's en methoden, maar praktische relevantie onder druk. *Tijdschrift voor Arbeidsvraagstukken*, 25, 318-335.
- Sternberg, R.J. & Lubart, T.I. (1999). The concept of creativity: Prospects and paradigms. In R.J. Sternberg (Ed.), *Handbook of creativity* (pp. 3-15). Cambridge, UK: Cambridge University Press.
- Thorndike, R.L. (1949). *Personnel selection*. New York: Wiley.
- Van Oudenhoven, J.P.L.M., Blank, A., Leemhuis, F., Pomp, M. & Sluis, A.F. (2008). *Nederland deugt*. Assen, Nederland: Van Gorcum.
- Wresniewski, A. & Dutton, J.E. (2001). Crafting a job: Revisioning employees as active crafters of their work. *Academy of Management Review*, 26, 179-201.

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.

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 functierelevantie 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 functierelevantie in deze businessgerelateerde vaardigheidstesten wordt bevorderd door bij de itempresentatie gebruik te maken van realistische grafieken, tabellen of functiespecifieke tekstpassages. Hatstrup, 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 scores op realisme en functie-relevantie zijn 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 persoonlijkheidstrekk 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 persoonlijkheidstrekk '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 multimediamaatstaf. 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 vooralsnog niet gebeurd.

Hoe kunnen organisaties de duistere kant van sollicitanten inschatten?

In de laatste tien jaar vormt het onderzoek naar de onderliggende structuur van het prestatiedomein een van de belangrijkste ontwikkelingen bij personeelsselectie. 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 psychologische 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 ‘persoonlijheidsgebaseerde 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 bestelen van zijn werkgever de collega’s moet verklappen die hem hebben geholpen bij de diefstal?’

De tweede categorie, met name de ‘persoonlijheidsgebaseerde integriteitstesten’, bevragen de eerlijkheid van de sollicitant minder onoverholend 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.

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 onderzoeks aandacht 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 prestatie meting (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 vervolledigen 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 typ patroon.

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.

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 wegen 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.

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 opteren 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 functierelevant 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.

Samenvatting

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.

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.
- Hatrup, 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.
- Hatrup, 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., Sablynski, 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 racioethnic 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 ratee 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.

Werkverslaving, een begrip gemeten

Wilmar Schaufeli, Corine van Wijhe, Maria Peeters & Toon Taris*

Werkverslaving of workaholisme wordt opgevat als overmatig en dwangmatig werken. Gebaseerd op deze omschrijving wordt de DUWAS (Dutch Workaholism Scale) geïntroduceerd, een korte zelfbeoordelinglijst die bestaat uit twee gecorreleerde subschalen van ieder vijf items: excessief werken (EW) en compulsief werken (CW). Op basis van een nationale database (n = 11.060) zijn afkapwaarden berekend met behulp waarvan de mate van werkverslaving kan worden vastgesteld. Uit analyses (n = 1.166) blijkt dat met name de combinatie van hoge scores op zowel EW als CW ongunstig is. Geconcludeerd wordt dat de DUWAS een betrouwbaar en valide instrument is dat bijvoorbeeld kan worden ingezet als screeningsinstrument om werkverslaafden op te sporen.

1 Inleiding

Het is niet verbazingwekkend dat juist in het huidige tijdsgewricht, waarin de nadruk ligt op hard werken, presteren en excelleren, de aandacht voor werkverslaving lijkt toe te nemen. Ook het feit dat men tegenwoordig, dankzij de moderne informatie- en communicatietechnologie, bijna overal en altijd kan werken, draagt daar aan bij. Bovendien loont hard werken. Zo bleek uit een enquête onder 20.000 lezers van *Intermediair* dat diegenen die tenminste 50 uur per week werkten, na 15 jaar gemiddeld ruim 50% méér verdienden dan diegenen die minder tijd aan hun werk besteedden (De Vos, 2003).

Alhoewel de term 'workaholisme' al ruim vier decennia geleden werd geïntroduceerd (Oates, 1968), duurde het tot de afgelopen eeuwwisseling voor er schot in het aantal wetenschappelijke publicaties over werkverslaving kwam. Zo telden Taris en Schaufeli (2007) tussen 2001 en 2006 in totaal 88 publicaties in *PsycLIT*, ongeveer evenveel als de 96 in de drie decennia daarvoor. Ter vergelijking: over burn-out zijn tussen 2001 en 2006 1.229 publicaties verschenen, ongeveer 14 keer zoveel dus. Er is geopperd dat dit relatief geringe aantal wetenschappelijke publicaties over workaholisme te maken heeft met het ontbreken van een algemeen geaccepteerde definitie en een daarop gebaseerd meetinstrument (McMillan & O'Driscoll, 2006). In dit artikel proberen we hiertoe een aanzet te geven door, op basis van een definitie die gebaseerd is op de meest gebruikte omschrijvingen van workaholisme, een korte zelfbeoordelingvragenlijst te intro-

* Universiteit Utrecht, Arbeids- en Organisationspsychologie

duceren. Deze vragenlijst – de DUTch WorkAholism Scale (DUWAS) – is inmiddels door ruim 10.000 Nederlandse werknemers ingevuld. Het onderhavige artikel beschrijft de constructie en de psychometrische evaluatie van de DUWAS, alsmede de normering van deze vragenlijst.

1.1 Het begrip workaholisme

Workaholisme is een woord met een knipoog. De geestelijke vader van de term, de Amerikaanse dominee en psychologieprofessor Wayne E. Oates (1971, p. 13), beschouwde dit neologisme als een ‘... semihumorous word for addiction to work’. Zoals wel vaker voorkomt bij psychologen was hijzelf slachtoffer van het verschijnsel dat hij beschreef. Uiteraard waren er ook mensen verslaafd aan hun werk voordat de term workaholisme bekend werd. Een voorbeeld van zo’n workaholic *avant la lettre* is de Franse schrijver Gustave Flaubert die in 1852 in een brief aan zijn maîtresse schreef: ‘I love my work with a frenetic and perverted love’ (geciteerd in Unwin, 2004, p. 10). Echter, pas na de publicatie van Oates’ bestseller in 1971 – met als titel *Confessions of a workaholic* – werd het verschijnsel wereldwijd bekend, eerst bij het grote publiek en daarna in de wetenschap.

Oates omschrijft workaholisme als ‘...the compulsion or the uncontrollable need to work incessantly’ (1971, p. 11). Er zijn hier dus twee wezenlijke elementen in het geding – *dwangmatig* en *onophoudelijk* werken –, die respectievelijk als de cognitieve en de gedragsmatige component van workaholisme gezien kunnen worden. Veel tijd besteden aan het werk is weliswaar een noodzakelijke maar geen voldoende reden om van *werkverslaving* te spreken (Machlowitz, 1980). Immers, men kan ook hard werken om veel geld te verdienen, om carrière te maken, om een ongelukkig huwelijk te ontvluchten, of omdat de leidinggevende het wil en de collega’s het ook doen. Daarmee is er nog geen sprake van een *werkverslaving* (Snir & Zohar, 2006). Die komt pas om de hoek kijken wanneer dit harde werken voortvloeit uit een niet te onderdrukken innerlijke drang. Met andere woorden, wanneer werken tot een obsessie wordt. *Werkverslaafden* voelen zich nutteloos, schuldig, geïrriteerd en gespannen wanneer ze niet kunnen werken, maar eenmaal aan het werk verdwijnen deze onlustgevoelens grotendeels (Porter, 1996; Scott, Moore & Miceli, 1997). Alhoewel er in de afgelopen drie decennia veel verschillende definities van workaholisme zijn verschenen loopt de oorspronkelijke omschrijving van Oates er als een rode draad doorheen. In een synthese van eerder onderzoek en theorievorming stellen Scott et al. (1997) dat in ieder geval aan twee eisen voldaan moet zijn om te kunnen spreken van *werkverslaving*. De eerste daarvan is dat *werkverslaafden* veel tijd besteden aan hun werk als zij daartoe de kans krijgen, hetgeen leidt tot verarming van hun sociale leven en verwaarlozing van allerlei andere (ontspannende) activiteiten. *Werkverslaafden* zijn dus *excessief harde* werkers. Het tweede element is dat *werkverslaafden* bijna altijd aan hun werk denken, dus ook buiten het werk in hun vrije tijd. *Werkverslaafden* zijn dus ook *dwangmatige* werkers.

Ook meer recente analyses van definities van workaholisme komen tot de conclusie dat *excessief hard* werken en een niet te stuiten innerlijke drang de kern uitmaken van *werkverslaving* (McMillan & O’Driscoll, 2006; Ng, Sorensen & Feldman, 2007). Dat onze bovenstaande definitie van workaholisme eveneens

overeenkomt met de opvatting die leken daarover hebben, wordt geïllustreerd door het onderzoek van McMillan en O'Driscoll (2006). Zij vroegen werknemers, hun partners en hun collega's hoe zij een workaholic zouden beschrijven. Na inhoudsanalyse van de gegeven antwoorden bleken de volgende twee elementen het meest te zijn genoemd: 'de grote hoeveelheid tijd die aan werken en daaraan denken wordt besteed' (39%) en 'een dwangmatige persoonlijkheidsstijl' (22%). De conclusie luidt dat de beide elementen die wij in navolging van Wayne E. Oates aan workaholisme onderscheiden – excessief hard werken en compulsief werken – als kernelementen worden teruggevonden in de meeste andere academische definities, alsmede in de opvatting die leken hebben over werkverslaving.

1.2 Het meten van werkverslaving

Overeenkomstig onze definitie van workaholisme meten we dit concept met behulp van twee schalen: excessief werken en compulsief werken. Ons uitgangspunt daarbij wordt gevormd door twee reeds bestaande schalen. Voor het meten van excessief werken wordt de *Compulsive Tendencies*-schaal van de *Work Addiction Risk Test* (WART; Robinson, 1999) gebruikt. De naam van deze schaal is enigszins misleidend omdat zeven van de negen items betrekking hebben op hard werken zonder daarbij te verwijzen naar de onderliggende, compulsieve motivatie. Vandaar dat we deze schaal bestaande uit zeven items herbenoemen tot Excessief Werken (EW). De twee overige items verwijzen naar het onvermogen om na werktijd te ontspannen en het zich schuldig voelen wanneer men niet aan het werk is. Een recente Nederlandse studie die met behulp van internet is uitgevoerd, geeft aan dat deze beide items op een aparte factor laden die compulsief werken meet (Schaufeli, Taris & Bakker, 2006). Voorts bleek uit een Nederlandse studie bij drie onafhankelijke steekproeven dat: (1) de score op de EW-schaal zeer hoog correleert met de totaalscore op de 15 items van de WART en (2) de correlaties van de EW-schaal en de WART met een aantal nevenvariabelen nagenoeg identiek zijn (Taris, Schaufeli & Verhoeven, 2005). Met andere woorden, de EW-schaal levert praktisch dezelfde informatie op als de complete WART, oftewel de EW-schaal is equivalent met de langere WART.

Voor het meten van compulsief werken wordt de *Drive*-schaal van de *Workaholism Battery* (WorkBat; Spence & Robbins, 1992) gebruikt. Alle acht items van deze schaal hebben betrekking op het dwangmatige karakter van het extreem harde werken van workaholics. Met het oog op onze definitie is deze schaal dan ook herbenoemd tot Compulsief Werken (CW). De interne consistentie van de CW-schaal blijkt voldoende te zijn in steekproeven uit verschillende landen, zoals Nieuw-Zeeland ($\alpha = .75$, McMillan, Brady, O'Driscoll & Marsh, 2002), Noorwegen ($\alpha = .80$, Burke & Matthiesen, 2004) en Japan ($\alpha = .70$, Kanai, Wakabayashi & Fling, 1996).

De keuze voor beide bovengenoemde schalen is dus gebaseerd op een gedegen analyse van de grootste gemene deler van de meeste definities van workaholisme. Voor een uitgebreide verantwoording van deze analyse en van de toegevoegde waarde van de combinatie van de EW- en de CW-schaal ten opzichte van andere bestaande schalen (Mudrack & Naughton, 2001; Robinson, 1999; Spence & Robbins, 1992) wordt verwezen naar Schaufeli, Taris en Shimazu (2009).

Inmiddels zijn enkele studies in Nederland uitgevoerd waarin *zowel* de EW- als de CW-schaal zijn opgenomen als indicatoren van workaholisme. Zo bevestigde de studie van Schaufeli et al. (2006) dat de twee eerder genoemde atypische EW-items laden op de CW-factor in plaats van de EW-factor. Nadat de twee EW-items aan de CW-schaal werden toegevoegd, verbeterde de interne consistentie zowel de EW- als de CW-schaal. Een andere studie onder managers toonde aan dat beide schalen sterk met elkaar samenhangen ($r = .59$), maar elkaar desalniettemin aanvullen; EW hangt vooral samen met kwantitatieve belasting zoals overwerk en werkdruk, terwijl CW vooral samenhangt met psychische en somatische klachten en gespannen sociale relaties (Schaufeli, Taris & Van Rhenen, 2008b).

1.3 De huidige studie

De huidige studie is bedoeld om een korte, betrouwbare en valide Nederlandse zelfbeoordelingvragenlijst voor workaholisme – de Dutch Workaholisme Scale (DUWAS) – te introduceren als hulpmiddel voor de praktijk. Het eerste doel van het onderzoek is dan ook het *ontwikkelen* van een korte vragenlijst met een goede factoriële validiteit. Dat wil zeggen, met twee positief samenhangende subschalen – excessief en compulsief werken – waarvan de items niet met elkaar overlappen. Uitgangspunt daarbij zijn de twee reeds bestaande Amerikaanse schalen die hierboven kort zijn besproken. Voor de ontwikkeling van de schaal wordt gebruik gemaakt van een aanzienlijke steekproef uit de Nederlandse beroepsbevolking. Een tweede doel is het *normeren* van de DUWAS. Dat wil zeggen, het bepalen van statistische afkappunten op basis waarvan workaholisme-scores als ‘(zeer) laag’, ‘gemiddeld’ dan wel ‘(zeer) hoog’ kunnen worden gekwalificeerd. Het derde doel is het *valideren* van de beide subschalen van de DUWAS door groepen werknemers die geselecteerd zijn op basis van hun hoge, dan wel lagere score op bovengenoemde schalen te vergelijken op een aantal nevenvariabelen. Omdat workaholisme door ons is gedefinieerd als de combinatie van excessief en compulsief werken willen we aantonen dat diegenen die een hoge score hebben op beide schalen van de DUWAS (‘workaholics’) het meest ongunstig scoren op een aantal nevenvariabelen in vergelijking tot andere werknemers, die ofwel alleen hard werken, ofwel alleen dwangmatig werken, ofwel noch hard, noch dwangmatig werken. De nevenvariabelen zijn geclusterd rondom de drie thema’s: werkomgevingskenmerken, persoonlijkheid en werkuitkomsten.

Werkomgevingskenmerken. Diverse kenmerken van de werkomgeving alsmede van de sociale omgeving buiten het werk zijn in verband gebracht met workaholisme. Daarbij gaat het zowel om potentiële stressoren als om energiebronnen. Zo is de positieve relatie tussen workaholisme en werkdruk in eerder onderzoek aangetoond; werkverslaafden zetten immers zichzelf onder druk door werk op te zoeken (Johnstone & Johnston, 2005). Het is evident dat het familieleven van een werkverslaafde in het gedrang komt door diens preoccupatie met het werk. Het wekt dan ook geen verbazing dat Bonebright, Clay en Ankenmann (2000) en Taris, Schaufeli en Verhoeven (2005) een positief verband vonden tussen werkverslaving en werk-thuis interferentie.

Het is voorts aannemelijk dat workaholisme negatief gerelateerd is aan energiebronnen. Zo kan de fanatieke en streberige werkhouding van workaholics belastend zijn voor diegenen met wie zij samenwerken (Porter, 2001) of samenleven (Bakker, Demerouti & Burke, 2009; Robinson, 1996). Ten aanzien van collega's kunnen workaholics bijvoorbeeld controlerend zijn (Mudrack & Naughton, 2001; Porter, 2004). Workaholics zijn bovendien weinig emotioneel betrokken bij hun thuissituatie (Robinson & Post, 1995). De partners en kinderen van workaholics moeten leven met een fysiek en emotioneel afwezige respectievelijke partner of ouder, hetgeen eenzaamheid en gevoelens van afwijzing (Robinson, 1998c), maar ook de ontwikkeling van psychologische problemen tot gevolg kan hebben (Robinson, 1998a). Workaholics stellen zich bovendien controlerend en perfectionistisch op in de relatie met hun familieleden (Robinson, 1998b). Als gevolg daarvan wordt de sociale steun van anderen op het werk (collega's en leidinggeevenden) alsmede in de privé-situatie ondermijnd. Uit het bovenstaande volgt: *Hypothese 1: Workaholics scoren het hoogst op de stressoren werkdruk en werk-thuis interferentie in vergelijking met andere groepen werknemers.*

Hypothese 2: Workaholics scoren het laagst op de energiebronnen sociale steun op het werk (collega's en leidinggeevenden) en in de privé-situatie (partner, familie en vrienden) in vergelijking met andere groepen werknemers.

Hypothese 3: Workaholics scoren het hoogst op overwerk en werken in het weekend in vergelijking met andere groepen werknemers.

De andere groepen werknemers hebben betrekking op respectievelijk 'harde werkers' (die hoog op EW en laag op CW scoren), 'dwangmatige werkers' (die hoog op CW en laag op EW scoren) en 'niet-werkverslaafden' (die laag op zowel EW als CW scoren). Vanzelfsprekend scoren workaholics zowel hoog op EW als op CW.

Persoonlijkheid. Onderzoek laat zien dat workaholisme gerelateerd is aan neuroticisme (Burke, Matthiesen & Pallesen, 2006a) en perfectionisme (Clark, Lelchook & Taylor, 2010; Spence & Robbins, 1992; Taris, Schaufeli & Van Beek, 2010). Overigens zijn neuroticisme en perfectionisme – opgevat als een dispositie om alles wat niet perfect is als onacceptabel te beschouwen – onderling nauw met elkaar verwant (Stoerber, Otto & Dahlbert, 2009). Het is aannemelijk dat het stellen van hoge eisen aan zichzelf, een kenmerk van perfectionisme dat wellicht voortkomt uit een fundamenteel gevoel van angst en onzekerheid (neuroticisme), tot gevolg heeft dat workaholics zo hard en dwangmatig werken. Ook het feit dat workaholics grote moeite hebben met het delegeren van taken naar anderen past in dit beeld (Killinger, 2006).

Hypothese 4: Workaholics scoren het hoogst op neuroticisme en op perfectionisme in vergelijking met andere groepen werknemers.

Werkuitkomsten. Verschillende onderzoeken laten zien dat de componenten van werkverslaving, en dan vooral het compulsieve aspect, negatief gerelateerd zijn aan bijvoorbeeld arbeidstevredenheid (Burke, Matthiesen & Pallesen, 2006b). Het is bovendien plausibel dat de betrokkenheid bij de organisatie positief samenhangt met workaholisme omdat workaholics gekenmerkt worden door een sterk plichtsbesef en dito verantwoordelijkheidsgevoel met betrekking tot de organisa-

tie waarvoor ze werken (Scott *et al.*, 1997). Voorts wordt op basis van eerder onderzoek (Burke, 2000) voorspeld dat workaholisme samenhangt met een verminderd welbevinden omdat workaholics zichzelf onvoldoende de kans geven om te herstellen van hun inspanningen, waardoor ze een verhoogde kans lopen op burn-out (Maslach, 1986; Andreassen, Ursin & Eriksen, 2007). Het tegenovergestelde van burn-out is bevlogenheid, een positieve, affectief-cognitieve toestand van opperste voldoening die wordt gekenmerkt door vitaliteit, toewijding en absorptie (Schaufeli & Bakker, 2001). Gegeven de positieve relatie met burn-out, wordt verwacht dat workaholisme negatief samenhangt met bevlogenheid (Schaufeli *et al.*, 2008a).

Hypothese 5: Workaholics scoren het laagst op arbeidssatisfactie en het hoogst op organisatiebetrokkenheid in vergelijking met andere groepen werknemers.

Hypothese 6: Workaholics scoren het hoogst op burn-out in vergelijking met andere groepen werknemers.

Hypothese 7: Workaholics scoren het laagst op bevlogenheid in vergelijking met andere groepen werknemers.

2. Methode

2.1 Respondenten

Voor de ontwikkeling en de normering van de DUWAS, respectievelijk het eerste en tweede doel van het huidige onderzoek, is gebruik gemaakt van een grootschalige steekproef van de Nederlandse beroepsbevolking ($n = 11.060$) van 16 jaar en ouder. De steekproef maakt deel uit van een internationaal databestand dat is samengesteld uit 15 verschillende onderzoeken. De steekproef bestaat voor circa de helft (45%) uit vrouwen. De gemiddelde leeftijd is 38 jaar ($sd = 10,39$). Deze beide kenmerken komen globaal overeen met die van de Nederlandse beroepsbevolking (51% vrouw, gemiddelde leeftijd bevolking 40 jaar, Centraal Bureau voor de Statistiek (CBS, 2009). Circa een derde van de steekproef is werkzaam in de gezondheidszorg en daarmee is deze sector enigszins oververtegenwoordigd. De zakelijke dienstverlening en de communicatiesector omvatten respectievelijk 14% en 10% van de steekproef en zijn daarmee enigszins ondervertegenwoordigd.

Voor de verdere validatie door middel van het analyseren van de combinatie van EW en CW, het derde doel van het huidige onderzoek, zijn diegenen uit de eerdergenoemde database geselecteerd ($n = 1.166$) waarvan een score op de dezelfde nevenvariabelen aanwezig was. Deze specifieke onderzoeksgroep omvat drie onafhankelijke steekproeven. Steekproef 1 en 2 bestaan uit werknemers van een Nederlands telecombedrijf (respectievelijk $n = 338$ leidinggevend en $n = 455$ werknemers). Steekproef 3 bestaat uit medewerkers van een adviesbureau ($n = 373$). Het onderzoek werd in alle gevallen uitgevoerd in het kader van een werkbelevingsenquête. Het overgrote deel van de geselecteerden was man (78%). De gemiddelde leeftijd van de werknemers bedroeg 39 jaar ($sd = 9,1$), variërend van 18 tot 62 jaar.

2.2 Instrumenten

2.2.1 Workaholisme

Workaholisme is gemeten met de schalen Excessief Werken (EW) en Compulsief Werken (CW) (zie tabel 1). De ontwikkeling van beide schalen wordt verderop uitvoerig beschreven. Alle items zijn gescoord op een 4-puntsschaal (1 = 'nooit', 4 = 'altijd').

2.2.2 Omgevingskenmerken

Omgevingskenmerken zijn onderverdeeld in potentiële stress- en energiebronnen in navolging van het Job Demands-Resources model (Bakker & Demerouti, 2007; Schaufeli, Bakker & Van Rhenen, 2009). Stress- en energiebronnen zijn gemeten met (verkorte) schalen die gebaseerd zijn op de Vragenlijst voor de Beleving en Beoordeling van Arbeid (VBBA; Van Veldhoven, De Jonge, Broersen, Kompier & Meijman, 2002) en die ook in ander onderzoek zijn gebruikt (zie bijvoorbeeld Schaufeli et al., 2008a-b, 2009). Alle items zijn gescoord op een 5-puntsschaal (1 = 'nooit', 5 = 'altijd'). De interne consistenties van alle schalen (Cronbach's α) zijn vermeld in tabel 2.

Stressbronnen zijn gemeten met de schalen Werkdruk (5 items, zoals 'Werkt u onder tijdsdruk?') en Werk-thuis interferentie (7 items, zoals 'Hoe vaak komt het voor dat u moeilijk aan de verplichtingen thuis kunt voldoen omdat u in gedachten steeds met het werk bezig bent?').

Energiebronnen zijn gemeten met schalen Steun van de leidinggevende (3 items, zoals 'Kunt u als dat nodig is uw directe leidinggevende om hulp vragen?'), Steun van de collega's (3 items, zoals 'Kunt u als dat nodig is collega's om hulp vragen?') en Steun uit de privésituatie (3 items, zoals 'Wordt u thuis werk uit handen genomen als u het op het werk drukker heeft?'). Laatstgenoemde schaal is afgeleid van de twee voorgaande schalen.

Overwerk is gemeten door te vragen naar werken in het weekend: 'Hoe vaak werkt u tijdens het weekend?' (1 = '(bijna) nooit', 4 = '(bijna) altijd'). Tevens is het percentage overwerk berekend door de overwerkuren te delen door de contracturen en te vermenigvuldigen met 100. De aldus verkregen overwerkratio, het percentage overwerk per week, bedroeg 20,8% ($sd = 17,3$). Het gemiddeld aantal contracturen was 38,3 ($sd = 3,3$) per week en het gemiddelde aantal overuren bedroeg 8,2 ($sd = 7,6$).

2.2.3 Persoonlijkheid

Perfectionisme is gemeten met een zelf geconstrueerde schaal (10 items, zoals 'Ik ben niet snel tevreden met het resultaat van mijn werk' en 'Ik haal vaak de laatste foutjes uit het werk van mijn collega's') en wordt gescoord op een 5-puntsschaal (1 = 'helemaal mee oneens', 5 = 'helemaal mee eens'). Het persoonlijkheidskenmerk neuroticisme is gemeten met de gelijknamige schaal Neuroticisme (12 items, zoals 'Soms voel ik me volkomen waardeloos') afkomstig uit de Nederlandse versie (Hoekstra, Ormel & De Fruyt, 1996) van de NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992), gebruikmakend van een 5-puntsschaal (1 = 'helemaal oneens', 5 = 'helemaal eens').

2.2.4 Werkuitskomsten

Uitkomsten zijn gemeten met twee zelfgeconstrueerde schalen (zie Schaufeli et al., 2008a-b, 2009) die gebaseerd zijn op de VBBA (Van Veldhoven et al., 2002): *Arbeidssatisfactie* (3 items, zoals 'Ik ben tevreden met mijn huidige werk') en *Organisatiebetrokkenheid* (5 items, zoals: 'Ik voel me thuis bij <naam organisatie>'). Alle items zijn gescoord op een 5-puntsschaal (1 = 'helemaal mee oneens', 5 = 'helemaal mee eens').

Burn-out is gemeten met twee van de drie subschalen van de Utrechtse Burn-out Schaal (UBOS) (Schaufeli & Van Dierendonck, 2000), de Nederlandse versie van de Maslach Burnout Inventory-General Survey (Schaufeli, Leiter, Maslach & Jackson, 1996). De twee kerndimensies zijn Uitputting (5 items, zoals 'Ik voel me mentaal uitgeput door mijn werk') en Distantie (4 items, zoals 'Ik merk dat ik te veel afstand heb gekregen van mijn werk'). Alle items zijn gescoord op een 7-puntsschaal (0 = 'nooit', 6 = 'altijd').

Bevlogenheid is gemeten met de Utrechtse Bevlogenheidschaal (UBES, Schaufeli, Salanova, González-Romá & Bakker, 2002). Dit meetinstrument bestaat uit drie subschalen, te weten Vitaliteit (6 items, zoals 'Op mijn werk bruis ik van energie'), Toewijding (5 items, zoals 'Ik ben trots op het werk dat ik doe') en Absorptie (6 items, zoals 'Ik ga helemaal op in mijn werk'). De items zijn gescoord op een 7-puntsschaal (0 = 'nooit', 6 = 'altijd').

3 Resultaten

3.1 De ontwikkeling van de DUWAS

In verband met mogelijke kanskapitalisatie is de database willekeurig gesplitst in twee steekproeven. De eerste, 'exploratieve' steekproef werd gebruikt om de DUWAS te ontwikkelen ($n = 5.492$), terwijl de tweede, 'confirmatieve' steekproef werd gebruikt om resultaten te kruisvalideren ($n = 5.498$).

Een exploratieve factoranalyse met behulp van de principale componentenmethode gevolgd door varimaxrotatie over de 8 items van de Excessief Werken (EW)-schaal en de 8 items van de Compulsief Werken (CW)-schaal, leverde drie factoren op met eigenwaarden groter dan één. De totale hoeveelheid verklaarde variantie bedroeg 54,3%. Alle EW-items met uitzondering van 'Ik voel me schuldig als ik niet aan het werk ben' laadden 0,44 of hoger op de eerste factor. Twee items die hoger dan 0,50 op de eerste factor laadden hadden eveneens een lading hoger dan 0,30 op een van beide andere factoren. De CW-items waren daarentegen verspreid over de tweede en de derde factor, waarbij drie min of meer gelijke dubbelladingen voorkwamen. De hoogste ladingen op de tweede factor hadden te maken met 'zich schuldig voelen' en 'niet kunnen ontspannen', terwijl de hoogste ladingen op de derde factor vooral te maken hadden met een 'sterke drang om te werken'. Na verwijdering van het 'verkeerd' ladende EW-item en de vijf dubbelladingen bleven er 10 items over, 5 voor ieder van de twee schalen. Zoals blijkt uit tabel 1 resulteerde een tweede exploratieve factoranalyse over de 10 overgebleven items in een fraaie *simple solution* waarbij alle items hoger dan 0,50 laadden op de bedoelde factor (vet gedrukt) en lager dan 0,30 op de andere factor. Op Factor 1

Tabel 1 Factorladingen van de verkorte Dutch Workaholism Scale (DUWAS) na varimax-rotatie (exploratieve steekproef, $n = 5,492$).

Item		Factor 1	Factor 2	h^2
		CW	EW	
1	Ik heb haast en werk tegen deadlines aan	0,15	0,75	0,59
2	Ik werk door terwijl mijn collega's al naar huis zijn	0,24	0,62	0,44
3	Ik vind het belangrijk om hard te werken, zelfs als ik eigenlijk geen plezier heb in mijn bezigheden	0,81	0,07	0,66
4	Ik ben druk en heb veel ijzers tegelijk in het vuur	0,10	0,83	0,70
5	Ik heb het gevoel dat iets in mijzelf me dwingt hard te werken	0,77	0,17	0,62
6	Ik besteed meer tijd aan mijn werk dan aan mijn vrienden, hobby's, of andere vrijetijdsactiviteiten	0,24	0,54	0,35
7	Ik voel me verplicht hard te werken, ook al vind ik dat niet altijd prettig	0,79	0,18	0,64
8	Ik ben met meerdere dingen tegelijk bezig, ik schrijf bijvoorbeeld een memo terwijl ik eet en met iemand telefoneer	0,16	0,69	0,50
9	Ik voel me schuldig als ik vrij neem van mijn werk	0,61	0,25	0,44
10	Ik vind het moeilijk om me te ontspannen als ik niet aan het werk ben	0,55	0,29	0,40

NB. EW = Excessief werken; CW = Compulsief werken.

laadden alle CW-items en op Factor 2 alle EW-items. De correlatie tussen beide schalen bedroeg 0,49 ($p < 0,001$). De correlatie van de verkorte CW-schaal met de oorspronkelijke schaal was 0,95 ($p < 0,001$) en die van de verkorte EW-schaal met de oorspronkelijke schaal was 0,92 ($p < 0,001$). Met andere woorden, de verkorte schalen waren nagenoeg identiek aan de oorspronkelijke versies. De waarden van Cronbach's α waren met respectievelijk 0,78 en 0,75 voor CW en EW ruim voldoende (Nunnally & Bernstein, 1994).

Vervolgens is ter kruisvalidatie in de tweede steekproef een confirmatieve factoranalyse uitgevoerd met behulp van het computerprogramma AMOS (Arbuckle, 2003). Als eerste werd een model getoetst met slechts één latente workaholismefactor waarop alle 10 items laadden, gevolgd door een twee-factormodel met een aparte EW- en CW-factor. Het één-factormodel paste niet goed bij de data ($\chi^2(35) = 4135,81$; GFI = 0,84; AGFI = 0,75; RMSEA = 0,15; NFI = 0,75; CFI = 0,74), terwijl de passing van het twee-factormodel goed was ($\chi^2(45) = 1471,33$; GFI = 0,95; AGFI = 0,91; RMSEA = 0,08; NFI = 0,91; CFI = 0,91). In het laatste geval voldeden alle passingsmaten aan de daarvoor geldende criteria (zie Byrne, 2009). Beide latente factoren correleerden .54 ($p < 0,001$). Voorts waren de correlaties tussen de oorspronkelijke en verkorte DUWAS-schalen identiek aan die van de exploratieve steekproef. De waarden van Cronbach's α verschilden met respectievelijk 0,79 en 0,75 voor CW en EW evenmin nauwelijks van die van de andere steekproef. Alles overziend kan geconcludeerd worden dat de kruisvalida-

tie succesvol verliep; de twee-factorstructuur zoals gevonden in de exploratieve steekproef paste goed bij de gegevens van de confirmatieve steekproef.

3.2 *Samenhang tussen workaholisme en nevenvariabelen*

Ter validatie van beide workaholismeschalen zijn de relaties met een aantal andere nevenvariabelen onderzocht. Ten eerste zijn Pearson product-moment correlaties berekend, waarbij bleek dat alle correlaties in de voorspelde richting waren. Zoals blijkt uit tabel 2 verschilt het patroon van correlaties van EW met dat van CW. EW hangt vooral samen met werkdruk, werk-thuis interferentie en overwerk, terwijl CW vooral samenhangt met neuroticisme, perfectionisme, burn-out (uitputting en cynisme) en een geringe mate van bevolegenheid (vitaliteit en toewijding). Met andere woorden, diegenen die excessief hard werken hebben het druk op het werk, werken veel uren en hebben problemen om hun werk te combineren met hun thuissituatie. Diegenen die dwangmatig werken hebben daarentegen meer psychische problemen.

3.3 *Normering*

EW en CW hangen zwak maar in het algemeen significant samen met geslacht ($t(10.983) = -4,39, p < 0,001$, respectievelijk $t(10.983) = 0,12, p = ns$) en leeftijd ($r = -0,03, p < .05$, respectievelijk $r = -0,08, p < 0,001$). Op een enkele uitzondering na worden bovendien geen relevante verschillen in workaholisme gevonden tussen beroepsgroepen of sectoren. Een groep die echter opvallend verschilt van de overige groepen is de kleine groep zelfstandige ondernemers ($n = 45$). Zij scoren in vergelijking met de andere beroepsgroepen significant hoger op EW en CW, $M = 2,88$ respectievelijk $M = 2,28$ tegen $M = 2,53$ respectievelijk $M = 2,00$ ($t(11.058) = 3,89, p < 0,001$, respectievelijk $t(11.058) = 2,86, p < 0,001$). Op grond van het feit dat – met uitzondering van ondernemers – geen of nauwelijks verschillen tussen beroepsgroepen of sectoren zijn gevonden is besloten om een uniforme normering te hanteren zonder daarbij rekening te houden met sekse, leeftijd, beroepsgroep en/of sector.

Op basis van de ruwe scores van de normpopulatie ($n = 11.060$) zijn vijf percentie-scores berekend voor de beide schalen van de DUWAS met behulp waarvan een classificatie in vijf categorieën kan worden gemaakt: ‘zeer laag’ ($\leq 5^e$ percentiel), ‘laag’ ($6^e - 25^e$ percentiel), ‘gemiddeld’ ($26^e - 75^e$ percentiel), ‘hoog’ ($76^e - 94^e$ percentiel), ‘zeer hoog’ ($\geq 95^e$ percentiel). Een dergelijke classificatie wordt ook gebruikt bij burn-out (Schaufeli & Van Dierendonck, 2000) en bevolegenheid (Schaufeli & Bakker, 2004). In tabel 3 staan de afkapwaarden die behoren bij de percentie-scores en die gebruikt kunnen worden voor de classificatie van het niveau van workaholisme.

3.4 *De combinatie van excessief en compulsief werken*

Op basis van de bovenstaande normering kunnen vier groepen worden onderscheiden: (1) ‘workaholics’ (hoog op zowel EW als CW); (2) ‘dwangmatige werkers’ (hoog op CW en gemiddeld of (zeer) laag op EW); (3) ‘harde werkers’ (hoog op EW en gemiddeld of (zeer) laag op CW); (4) ‘niet-werkverslaafde werkers’ (gemiddeld

Tabel 2 Gemiddelden (M), standaarddeviaties (SD), interne consistentie (Cronbach's α) en correlaties van de variabelen met workaholisme (n = 1,166).

Variabele	M	SD	α	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1. Workaholisme	2.47	0.42	.79	-																			
2. Excessief werken	2.80	0.49	.76	.82	-																		
3. Compulsief werken	2.15	0.53	.83	.85	.39	-																	
4. Werkdruk	2.98	0.73	.84	.50	.57	.29	-																
5. Werk-thuis interferentie	2.12	0.70	.90	.59	.56	.43	.51	-															
6. Steun privé	3.81	1.11	.87	-.10	.01	-.16	-.02	-.09	-														
7. Steun collega's	3.79	0.81	.87	-.23	-.14	-.24	-.16	-.21	.25	-													
8. Steun leidinggevende	3.79	0.89	.86	-.24	-.13	-.26	-.17	-.27	.21	.48	-												
9. Overwerk (%)	20.75	17.30	nvt	.25	.41	.03	.29	.33	.16	.02	.03	-											
10. Overwerk weekend	1.88	0.81	nvt	.29	.36	.13	.25	.39	.00	-.06	-.05	.34	-										
11. Neuroticisme	2.06	0.55	.84	.38	.13	.49	.12	.27	-.24	-.26	-.29	-.13	-.01	-									
12. Perfectionisme	3.29	0.46	.72	.34	.23	.33	.13	.20	-.05	-.06	.00	.05	.09	.17	-								
13. Arbeidsatisfactie	3.88	0.71	.90	-.25	-.11	-.29	-.19	-.22	.12	.40	.47	.10	.02	-.36	-.01	-							
14. Organisatiebetrokkenheid	3.30	0.70	.82	.10	.08	.08	.00	-.02	.04	.13	.17	.14	.09	-.05	.12	.31	-						
15. Uitputting	1.31	0.87	.85	.52	.36	.50	.39	.50	-.13	-.21	-.26	.05	.07	.49	.20	-.42	-.08	-					
16. Distantie	1.11	0.91	.78	.30	.17	.33	.25	.31	-.13	-.32	.43	-.09	.04	.38	.02	-.64	-.26	.52	-				
17. Vitaliteit	4.11	0.83	.86	-.08	.13	-.25	.02	-.07	.21	.30	.31	.28	.17	-.49	.08	.54	.29	-.39	-.45	-			
18. Toewijding	4.18	0.98	.92	-.14	.04	-.26	-.05	-.13	.19	.37	.42	.21	.11	-.42	.07	.73	.38	-.36	-.63	.77	-		
19. Absorptie	3.55	0.82	.74	.19	.25	.08	.14	.13	.12	.19	.22	.26	.20	-.13	.22	.42	.37	-.08	-.34	.68	.68	-	

NB. Correlaties van .06 en hoger zijn significant; $p < 0,05$

Tabel 3 Gemiddelden (*M*), standaarddeviaties (*SD*) en percentielscores (*p*) voor Excessief Werken (*EW*) en Compulsief Werken (*CW*) (*n* = 11,060).

	EW	CW
M	2,53	2,01
SD	0,60	0,64
Zeer laag (p5)	1,40	1,00
Laag (p25)	1,41 – 2,00	1,01 – 1,50
Gemiddeld (p50)	2,01 – 2,90	1,51 – 2,40
Hoog (p75)	2,91 – 3,40	2,41 – 3,00
Zeer hoog (p95)	3,41	3,01

of (zeer) laag op zowel EW als CW). Met behulp van ANOVA's is onderzocht hoe de groepen ten opzichte van elkaar scores ten aanzien van eerdergenoemde drie domeinen (zie tabel 4). In de onderstaande beschrijving van de resultaten zijn de scores van de 'workaholics' – conform de hypothesen – gecontrasteerd met de scores van de overige drie groepen.

Uit *post hoc*-analyse (zie tabel 4) blijkt dat juist workaholics zijn blootgesteld aan stressbronnen; ze ervaren meer werkdruk en werk-thuis interferentie dan de drie overige groepen (Hypothese 1 bevestigd). Ook beschikken workaholics over minder energiebronnen dan harde werkers (steun vanuit de privésituatie, van collega's en de leidinggevende) en niet-werkverslaafde werkers (steun van collega's en de leidinggevende). Ze verschillen echter niet van dwangmatige werknemers in de mate van ervaren steun (Hypothese 2 deels bevestigd). Verder maken workaholics meer overuren en werken ze vaker in het weekend dan dwangmatige werkers en niet-werkverslaafde werkers (Hypothese 3 deels bevestigd). Met betrekking tot persoonlijkheid (neuroticisme en perfectionisme) scoren workaholics ongunstiger dan de harde werkers en niet-werkverslaafden, maar gunstiger dan dwangmatige werkers (Hypothese 4 deels bevestigd). Ten aanzien van werk-uitkomsten scoren workaholics op arbeidssatisfactie in verhouding tot harde werkers en niet-werkverslaafde werknemers laag. Workaholics verschillen echter niet van de andere groepen ten aanzien van hun betrokkenheid bij de organisatie (Hypothese 5 deels bevestigd). Workaholics voelen zich in vergelijking met de overige drie groepen het meest uitgeput en ze scoren hoger op distantie dan de harde werkers en niet-werkverslaafde werknemers. Met andere woorden: workaholics hebben relatief veel burn-outklachten (Hypothese 6 grotendeels bevestigd). Ofschoon workaholics lager scoren op vitaliteit dan harde werkers, voelen ze zich energiekter dan dwangmatige werkers. Ze scoren lager op toewijding dan harde werkers en niet-werkverslaafde werkers, maar niet lager dan dwangmatige werkers. Tot slot voelen ze zich vaker geabsorbeerd door hun werk dan dwangmatige werkers en niet-werkverslaafde werkers, maar niet vaker dan harde werkers (Hypothese 7 deels bevestigd). Kort samengevat scoren werkverslaafden over het algemeen ongunstiger op omgevingskenmerken, persoonlijkheid en uitkomsten

in vergelijking met de overige drie groepen. Echter, in sommige gevallen zijn de scores van met name de dwangmatige werkers, vergelijkbaar of zelfs wat ongunstiger dan die van werkverslaafde werkers.

4. Discussie

In dit onderzoek is een nieuw Nederlandstalig instrument om workaholisme te meten geïntroduceerd; de Dutch WorkAholism Scale (DUWAS). Op basis van de gegevens van ruim 11.000 Nederlandse werknemers werd duidelijk dat werkverslaving betrouwbaar gemeten kan worden met twee relatief korte schalen van vijf items Excessief Werken (EW) en Compulsief Werken (CW). Beide schalen overlappen elkaar gedeeltelijk, dat wil zeggen hun gedeelde variantie is ongeveer 30%. Het feit dat het bij workaholisme gaat om de combinatie van excessief werken (de gedragsmatige component) en compulsief werken (de cognitieve component) wordt empirisch ondersteund doordat diegenen die hoog scores op beide schalen van de DUWAS, blootstaan aan de meest ongunstige omgevingskenmerken, de meest problematische persoonlijkheid hebben en de slechtste uitkomsten rapporteren. Met andere woorden, er kan dus geconcludeerd worden dat de DUWAS op betrouwbare en valide wijze de tweedimensionele omschrijving van workaholisme weerspiegelt, zoals deze elders is geformuleerd (Taris, Schaufeli & Shimazu, 2010; Schaufeli, Taris & Bakker, 2008a). De meerwaarde van de huidige studie ten opzichte van eerdergenoemde artikelen ligt in de ontwikkeling van normscores, welke een handvat kunnen bieden voor bedrijfsartsen en A&O-psychologen om workaholisme te onderscheiden. Het nut van de normscores wordt geïllustreerd door de bevinding dat werknemers met een combinatie van hoge scores op excessief en compulsief werken over het algemeen ongunstigere uitkomsten ervaren dan werknemers met sec hoge scores op excessief dan wel compulsief werken.

Er zijn echter ook enkele onverwachte resultaten gevonden die om nadere beschouwing vragen. Zo blijkt Hypothese 1 over de relatie tussen workaholisme en stressoren wel ondersteund te worden, maar Hypothese 2 over de relatie tussen workaholisme en energiebronnen slechts gedeeltelijk. Het opmerkelijke daarbij is dat workaholics niet verschillen van 'dwangmatige werkers' in ervaren steun vanuit hun omgeving. Workaholics verschillen ook niet van dwangmatige werkers in hun mate van perfectionisme en neuroticisme (Hypothese 4), arbeidssatisfactie (Hypothese 5), distantie (Hypothese 6) en toewijding (Hypothese 7) ten opzichte van hun werk. Bovendien scoren dwangmatige werkers zelfs ongunstiger dan workaholics op vitaliteit en absorptie (Hypothese 7). Blijkbaar hebben beide typen werknemers een enigszins overeenkomstig profiel wat betreft persoonlijkheidskenmerken en uitkomsten. Dit is niet verwonderlijk aangezien het compulsieve element de meest centrale en nadelige component is van workaholisme (Schaufeli et al., 2008a-b).

Het feit dat workaholics in verhouding hoog scores op absorptie is ook niet opzienbarend omdat workaholisme gekarakteriseerd wordt door het zich volledig onderdompelen in het werk (Porter, 1996) en door het onvermogen om het werk

Tabel 4 Gemiddelden (M) en standaarddeviaties (SD) op verschillende domeinen voor *Workaholics (WA)*, *Dwangmatige Werkers (DW)*, *Harde Werkers (HW)* en *Niet-Werkverslaafden (NW)*.

	WA n = 193 M (SD)	DW n = 96 M (SD)	HW n = 285 M (SD)	NW n = 583 M (SD)	F
Omgevingskenmerken					
Werkdruk	3,51 (0,64) ^{abc}	2,87 (0,64) ^{ad}	3,28 (0,67) ^{bde}	2,69 (0,64) ^{ce}	40,40
Werk-thuis interferentie	2,76 (0,77) ^{abc}	2,22 (0,59) ^{ad}	2,31 (0,61) ^{be}	1,79 (0,55) ^{cde}	44,41
Sociale steun privé	3,65 (1,21) ^a	3,33 (1,24) ^{bc}	3,95 (1,04) ^{ab}	3,86 (1,07) ^c	137,22
Sociale steun collega's	3,52 (0,82) ^{ab}	3,56 (0,87) ^c	3,76 (0,77) ^{ad}	3,92 (0,78) ^{bcd}	20,57
Sociale steun leidinggevende	3,49 (0,93) ^{ab}	3,49 (0,92) ^{cd}	3,83 (0,85) ^{ac}	3,91 (0,86) ^{bd}	9,39
Werken in het weekend	2,21 (0,87) ^{ab}	1,78 (0,70) ^{ac}	2,15 (0,83) ^{cd}	1,65 (0,72) ^{bd}	26,27
Overwerkcratio	26,27 (20,38) ^{ab}	13,97 (15,91) ^{ac}	28,10 (15,95) ^{cd}	16,31 (15,05) ^{bd}	26,68
Persoonlijkheid					
Neuroticisme	2,38 (0,54) ^{ab}	2,49 (0,49) ^{cd}	1,95 (0,49) ^{ac}	1,95 (0,52) ^{bd}	30,96
Perfectionisme	3,52 (0,47) ^{ab}	3,42 (0,44) ^c	3,31 (0,44) ^{ad}	3,19 (0,43) ^{bcd}	60,78
Uitkomsten					
Arbeidsatisfactie	3,56 (0,76) ^{ab}	3,63 (0,69) ^{cd}	3,95 (0,67) ^{ac}	3,99 (0,68) ^{bd}	23,40
Organisatiebetrokkenheid	3,40 (0,72)	3,39 (0,77)	3,28 (0,68)	3,25 (0,68)	2,62
Uitputting	2,07 (0,95) ^{abc}	1,76 (0,89) ^{ade}	1,27 (0,70) ^{bdf}	1,01 (0,72) ^{cef}	102,57
Distantie	1,53 (0,99) ^{ab}	1,58 (0,89) ^{cd}	1,05 (0,86) ^{ac}	0,93 (0,84) ^{bd}	33,06
Vitaliteit	3,99 (0,71) ^{ab}	3,57 (0,87) ^{acd}	4,39 (0,70) ^{bce}	4,09 (0,87) ^{de}	27,27
Toewijding	3,97 (0,96) ^{ab}	3,66 (1,00) ^{cd}	4,37 (0,90) ^{ac}	4,23 (0,98) ^{bd}	16,83
Absorptie	3,79 (0,80) ^{ab}	3,36 (0,89) ^{ac}	3,72 (0,79) ^{cd}	3,42 (0,81) ^{bd}	16,54

NB. Overeenkomstige subscripts indiceren significante verschillen tussen de gemiddelden; $p < 0,01$.

los te laten (McMillan et al., 2001). Een dergelijke omschrijving overlapt sterk met die van absorptie als ‘...being fully concentrated and deeply engrossed in one’s work, whereby time passes quickly and one has difficulties with detaching oneself from work’ (Schaufeli et al., 2002, p. 75). Met andere woorden, workaholisme en bevologenheid overlappen elkaar voor wat betreft het geheel en al opgaan in het werk (absorptie). Echter, op basis van de Self-Determination Theory (Gagné & Deci, 2005) kan vermoed worden dat de onderliggende motivatie fundamenteel verschilt: bij bevologenheid is er sprake van intrinsieke motivatie, terwijl er bij workaholisme sprake is van gecontroleerde, geïntrojecteerde motivatie (Van Beek, Taris & Schaufeli, 2010). Het laatste verwijst naar een motivationele regulatiemodus waarbij taak-extrinsieke elementen zoals behoud van zelfwaarde en sociale erkenning zijn geïnternaliseerd en het werkgedrag sturen (Koestner & Losier, 2004). Oftewel, er is dus sprake van plezier tegenover innerlijke dwang om te werken.

Het gegeven dat workaholics vitaler zijn dan dwangmatige werknemers is minder inzichtelijk. Men zou kunnen veronderstellen dat een bepaalde mate van vitaliteit een voorwaarde is om hard te kunnen werken. Dwangmatige werkers hebben wel de drang, maar niet de energie om hard te werken. Echter, workaholics ervaren wel meer uitputting dan de overige groepen, waaronder dwangmatige werkers. Dit zou kunnen betekenen dat voor dwangmatige werkers de lage mate van vitaliteit een beschermende factor zou kunnen zijn voor burn-out; hun lage energieniveau voorkomt dat ze gehoor geven aan hun dwang om hard te werken. Over de relatie tussen bevologenheid, burn-out en workaholisme is echter nog te weinig bekend om goed onderbouwde uitspraken te kunnen doen.

Een andere opvallende bevinding is dat Hypothese 3, waarin werd gesteld dat workaholics de meeste overuren maken, slechts deels wordt bevestigd. Harde werkers verschillen namelijk niet van workaholics in het aantal uren overwerk. Bij nader inzien is dit niet verwonderlijk, omdat wanneer iemand excessief werkt het niet direct hoeft te suggereren dat diegene werkverslaafd is. Men kan hard werken om veel verschillende redenen. Eén daarvan is dat een werknemer heel bevologen kan zijn in zijn of haar werk (Schaufeli et al., 2006). Het resultaat dat de groep harde werkers een hoge mate van vitaliteit rapporteert, onderstreept de redenering dat sommige werkers hard werken, omdat ze zich energiek voelen en zin hebben om aan het werk te zijn, zonder daarbij werkverslaafd te zijn.

Tot slot is het opmerkelijk dat workaholics ten aanzien van hun betrokkenheid bij de organisatie niet verschillen van de overige groepen. Blijkbaar voelen de workaholics zich niet in sterkere mate ‘een deel’ van de organisatie dan dat andere werkers dat doen. Voor workaholics is wellicht vooral hard werken belangrijk en tot op een bepaalde hoogte maakt het niet uit voor welke organisatie ze dat doen.

Beperkingen van de studie

Een evidente beperking is het feit dat uitsluitend zelfrapportages zijn gebruikt. Een interessante optie zou in dit verband zijn om de oordelen van anderen zoals collega’s, leidinggevenden en partners te relateren aan de scores op de DUWAS.

Burke en Ng (2007) vonden correlaties tussen zelfrapportage op de WorkBat (Spence & Robbins, 1972) en beoordeling door collega's die varieerden van 0,26 tot 0,57, al naar gelang de workaholisme-dimensie. Ten slotte zijn er geen instrumenten opgenomen die soortgelijke begrippen als workaholisme meten zoals 'obsessive passion' (Vallerand, Blanchard et al., 2003), 'overcommitment' (Siegrist, 1998) en Type A-persoonlijkheid (Day & Jreige, 2002). Hierdoor was het niet mogelijk om de convergente validiteit van de DUWAS vast te stellen. Ten slotte, perfectionisme is gemeten als een eendimensionaal construct. Uit onderzoek blijkt echter dat perfectionisme uit een adaptieve en maladaptieve dimensie bestaat, respectievelijk een hoge standaard voor jezelf hebben en bang zijn om fouten te maken (Frost, Marten, Lahart & Rosenblate, 1990). Daarbij rijst de vraag hoe deze dimensies van perfectionisme op verschillende wijze samenhangen met workaholisme. Een duidelijker onderscheid tussen deze vormen van perfectionisme biedt een beter inzicht in de risicofactoren voor workaholisme. Taris, Van Beek en Schaufeli (in druk), toonden recentelijk aan dat workaholisme met name samengaat met het maladaptieve element van perfectionisme en niet met de adaptieve component van perfectionisme.

Implicaties

De resultaten van ons onderzoek impliceren dat met de DUWAS op betrouwbare en valide wijze onderzoek kan worden gedaan naar werkverslaving in Nederland. Het huidige onderzoek laat bovendien zien dat de combinatie van excessief en compulsief werken ogenschijnlijk ongunstige gevolgen heeft. De huidige studie biedt verschillende handvatten om het werkverslavingsprobleem aan te pakken. Enerzijds leent de DUWAS zich vanwege haar theoretische verankering voor wetenschappelijk onderzoek naar oorzaken en gevolgen van workaholisme en naar relaties met andere, soortgelijke begrippen (Schaufeli et al., 2009).

Anderzijds kan de DUWAS in de praktijk worden ingezet als screeningsinstrument om werkverslavingsproblematiek tijdig te identificeren. De normgegevens kunnen bijvoorbeeld door arbodiensten in het kader van een Preventief Medisch Onderzoek (PMO) gebruikt worden om specifieke 'risicogevalen' op te sporen en de betreffende werknemers een geschikt advies te geven voor een vervolgtraject. Vroegtijdige signalering is van belang omdat het kan bijdragen aan een effectieve preventiebehandeling van workaholisme (Van Wijhe, Peeters & Schaufeli, 2010). Meer gedetailleerd onderzoek naar de sensitiviteit en de specificiteit van de DUWAS is echter wenselijk, met name met het oog op toepassing van de DUWAS voor individuele diagnostiek.

Voorts kan met behulp van de normscores het aantal werkverslaafden worden vastgesteld in teams, afdelingen, functiegroepen of in organisaties als geheel, bijvoorbeeld in het kader van een werkbelevingsonderzoek. Door de DUWAS op te nemen in zo'n doorgaans periodiek onderzoek, kan gericht HRM-beleid worden gevoerd om werkverslaving te voorkomen. Immers, zoals uit ons onderzoek blijkt kan een organisatie maar beter niet te veel werkverslaafden hebben; ze zijn minder tevreden over hun werk, ervaren ze meer burn-outklachten en zijn ze minder

bevlogen dan hun niet-werkverslaafde collega's. Het is dus van belang dat managers en andere leidinggevendenden op de hoogte zijn van de mate waarin hun werknemers werkverslaafd zijn.. De DUWAS kan daarbij dus worden ingezet.

Ten slotte kan de DUWAS, behalve als screeningsinstrument, ook gebruikt worden om het effect van interventies voor workaholisme te evalueren. Zo loopt er momenteel een onderzoek naar het effect van een e-coachingsmodule die door werknemers met (dreigende) werkverslavingsproblematiek gevolgd kan worden, en waarbij de DUWAS als criteriummaat wordt gebruikt.

Kortom, met de DUWAS beschikken we over een betrouwbaar en valide instrument dat zowel in de wetenschap als in de praktijk kan worden gebruikt om workaholisme te meten.

Praktijkbox

- Wat betekenen de resultaten voor de praktijk? Hoe te werken aan de screening en preventie van werkverslaving?
- Werkverslaving bestaat uit een combinatie van hard werken en dwangmatig werken. Hardwerkende werknemers zijn dus op zichzelf geen reden tot zorg, maar wanneer werknemers ook dwangmatig werken kan dit negatieve effecten hebben.
- De DUWAS kan als screeningsinstrument voor werkverslaving worden gebruikt, bijvoorbeeld bij Preventief Medisch Onderzoek (PMO) in het kader van de Arboret. Aldus kan de DUWAS bijdragen aan de vroegtijdige ontdekking of er sprake is van (een risico op) werkverslaving, hetgeen interventie vergemakkelijkt.
- Op basis van normscores kan de ernst van de werkverslavingsproblematiek in kaart worden gebracht. Met name wanneer een werknemer zeer hoog op beide schalen van de DUWAS scoort, is de kans op nadelige consequenties het grootst.
- De DUWAS kan worden ingezet voor evaluatie en effectmeting ten behoeve van interventies. Dit draagt bij aan een kwaliteitsverbetering van interventies, maar ook inzicht in de veranderbaarheid van werkverslaafd gedrag.
- Door de DUWAS op te nemen in werkbelevingsonderzoek kan gericht HRM-beleid gevoerd worden met het doel om werkverslaving te voorkomen.

Samenvatting

Workaholism or work addiction is defined as the uncontrollable need to work excessively hard. Based on this definition, the DUWAS (Dutch Workaholism Scale) is introduced, a brief two-dimensional self-report instrument that includes two correlated subscales of five items each: Working Excessively (WE) and Working Compulsively (WC). Using a national Dutch database ($n = 11,060$) cut-off scores were computed that allow a classification of levels of workaholism. Analyses ($n = 1,166$) reveal that particularly the combination of high scores on both WE

and CW is unfavorable. It is concluded that the DUWAS is a reliable and valid inventory that can, for instance, be used as a screening tool to identify workaholic employees.

Literatuur

- Andreassen, C.S., Ursin, H., & Eriksen, H.R. (2007). The relationship between strong motivation to work, 'workaholism', and health. *Psychology and Health, 22*, 615-629.
- Arbuckle, J.L. (2003). *AMOS 5.0 (Computer Software)*. Chicago: SPSS.
- Bakker, A.B. & Demerouti, E. (2007). The Job Demands-Resources model: State of the art. *Journal of Managerial Psychology, 22*, 309-328.
- Bakker, A.B., Demerouti, E., & Burke, R. (2009). Workaholism and relationship quality: A spillover-crossover perspective. *Journal of Occupational Health Psychology, 14*, 23-33.
- Bonebright, C.A., Clay, D.L., & Ankenmann, R.D. (2000). The relationship of workaholism with work-life conflict, life satisfaction, and purpose in life. *Journal of Counseling Psychology, 47*, 469-477.
- Burke, R. (2000). Workaholism in organizations: Psychological and physical well-being consequences. *Stress and Health, 16*, 11-16.
- Burke, R.J. & Koksal, H. (2002). Workaholism among a sample of Turkish managers and professionals: An exploratory study. *Psychological Reports, 91*, 60-68.
- Burke, R.J. & Matthiesen S.B. (2004). Workaholism among Norwegian journalists: Antecedents and consequences. *Stress & Health, 20*, 301-308.
- Burke, R.J. & Ng, E.S.W. (2007). Workaholic behaviors: Do colleagues agree? *International Journal of Stress Management, 14*, 312-320.
- Burke, R.J., Matthiesen, S.B., & Pallesen, S. (2006a). Personality correlates of workaholism. *Personality and Individual Differences, 40*, 1223-1233.
- Burke, R.J., Matthiesen, S.B., & Pallesen, S. (2006b). Workaholism, organizational life and well-being of Norwegian nursing staff. *Career Development International, 11*, 463-477.
- Burke, R.J., Richardsen, A.M., & Martinussen, M. (2002). Psychometric properties of Spence and Robbins' measures of workaholism components. *Psychological Reports, 91*, 1098-1104.
- Byrne, B. (2009). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2de editie). Hillsdale (NJ): Lawrence Erlbaum Associates.
- CBS (2009). Bevolking; kerncijfers naar diverse kenmerken. Centraal Bureau voor de Statistiek. Gedownload op 6 januari 2009 van <http://statline.cbs.nl/>
- Clark, M.A., Lelchook, A.M., & Taylor, M.L. (2010). Beyond the Big Five: How narcissism, perfectionism, and dispositional affect relate to workaholism. *Personality and Individual Differences, 48*, 786-791.
- Costa, P.T. & McCrae, R.R. (1992). *NEO-PI-R Professional Manual*. Odessa (FL): Psychological Assessment Resources.
- Day, A.L. & Jreige, S. (2002). Examining Type A behavior pattern to explain the relationship between job stressors and psychosocial outcomes. *Journal of Occupational Health Psychology, 7*, 109-120.
- De Vos, E. (2003). Workaholisme: De kick van een 80-urige werkweek. *Intermediair*, 17 april, 12-17.
- Frost, O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research, 14*, 449-468.
- Gagné, M. & Deci, E.L. (2005) Self-determination theory and work motivation. *Journal of Organizational Behavior, 26*, 331-362.

- Hoekstra, H.A., Ormel, J., & De Fruyt, F. (1996). *Handleiding NEO Persoonlijkeids-vragenlijsten NEO-PI- R en NEO-FFI*. Lisse: Swets & Zeitlinger.
- Johnstone, A. & Johnston, L. (2005). The relationship between organizational climate, occupational type, and workaholism. *New Zealand Journal of Psychology*, 34, 181-188.
- Kanai, A., Wakabayashi, M., & Fling, S. (1996). Workaholism among employees in Japanese corporations: An examination based on the Japanese version of the Workaholism Scales. *Japanese Psychological Research*, 38, 192-203.
- Killinger, B. (2006). The workaholic breakdown syndrome. In R.J. Burke (Ed.), *Research companion to working time and work addiction* (pp. 61-88). Cheltenham (UK): Edward Elgar.
- Koestner, R. & Losier, G.F. (2004). Distinguishing three ways of being highly motivated: A closer look at introjection, identification, and intrinsic motivation. In E.L. Deci & R.M. Ryan (Eds.), *Handbook of self-determination research* (pp. 101-122). Rochester (NY): University of Rochester Press.
- Machlowitz, M. (1980). *Workaholics. Living with them, working with them*. Reading (MA): Addison-Wesley.
- Maslach, C. (1986). *Stress, burnout and workaholism*. In R.R. Killberg, P.E. Nathan, & R.W. Thoreson (red.), *Professionals in distress: Issues, syndromes and solutions in psychology* (pp. 53-73). Washington (DC): American Psychological Association.
- McMillan, L.H.W. & O'Driscoll, M.P. (2006). Exploring new frontiers to generate an integrated definition of workaholism. In R. Burke (red.), *Research companion to working time and work addiction* (pp. 89-107). Cheltenham (UK): Edward Elgar.
- McMillan, L.H.W., Brady, E.C., O'Driscoll, M.P., & Marsh, N. (2002). A multifaceted validation study of Spence and Robbins' (1992) Workaholism Battery. *Journal of Occupational and Organizational Psychology*, 75, 357-368.
- McMillan, L.H.W., O'Driscoll, M.P., Marsh, N.V., & Brady, E.C. (2001). Understanding workaholism: Data synthesis, theoretical critique, and future design strategies. *International Journal of Stress Management*, 8, 69-91.
- Mudrack, P.E. (2004). Job involvement, obsessive-compulsive personality traits, and workaholic behavioral tendencies. *Journal of Organizational Change Management*, 7, 490-508.
- Mudrack, P.E. & Naughton, T.J. (2001). The assessment of workaholism as behavioral tendencies: Scale development and preliminary empirical testing. *International Journal of Stress Management*, 8, 93-111.
- Ng, T.W.H., Sorensen, K.L., & Feldman, D.C. (2007). Dimensions, antecedents, and consequences of workaholism: a conceptual integration and extension. *Journal of Organizational Behavior*, 28, 111-136.
- Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric theory* (3de editie). New York: McGraw-Hill.
- Oates, W. (1968). On being a 'workaholic' (a serious jest). *Pastoral Psychology*, 19, 16-20.
- Oates, W. (1971). *Confessions of a workaholic: The facts about work addiction*. New York: World Publishing Company.
- Porter, G. (1996). Organizational impact of workaholism: Suggestions for researching the negative outcomes of excessive work. *Journal of Occupational Health Psychology*, 1, 70-84.
- Porter, G. (2001). Workaholic tendencies and the high potential for stress among co-workers. *International Journal of Stress Management*, 8, 147-164.
- Porter, G. (2004). Work, work ethic, work excess. *Journal of Organizational Change Management*, 17, 424-439.

- Robinson, B.E. (1996). The relationship between work addiction and family functioning: Clinical implications for marriage and family therapists. *Journal of Family Psychotherapy*, 7, 13-39.
- Robinson, B.E. (1998a). *Chained to the desk: a guidebook for workaholics, their partners and children, and the clinicians who treat them*. New York: New York University Press.
- Robinson, B.E. (1998b). Spouses of workaholics: Clinical implications for psychotherapy. *Psychotherapy*, 35, 260-268.
- Robinson, B.E. (1998c). The workaholic family: A clinical perspective. *American Journal of Family Therapy*, 26, 63-73.
- Robinson, B.E. (1999). The Work Addiction Risk Test: Development of a tentative measure of workaholism. *Perceptual and Motor Skills*, 88, 199-210.
- Schaufeli, W. & Bakker, A. (2001). Werk en welbevinding. Naar een positieve benadering in de arbeids- en gezondheidspsychologie. *Gedrag & Organisatie*, 5, 229-253.
- Schaufeli, W.B. & Bakker, A.B. (2004). Bevoegenheid: Een begrip gemeten. *Gedrag & Organisatie*, 17, 89-112.
- Schaufeli, W.B., Bakker, A.B., & Van Rhenen, W. (2009). How changes in job demands and resources predict burnout, work engagement, and sickness absenteeism. *Journal of Organizational Behavior*, 30, 893-917.
- Schaufeli, W.B. & Van Dierendonck, D. (2000). *Handleiding van de Utrechtse Burnout Schaal (UBOS)*. Lisse: Swets & Zeitlinger.
- Schaufeli, W.B., Taris, T.W., & Bakker, A. (2006). Dr. Jekyll and Mr. Hyde: On the differences between work engagement and workaholism. In R. Burke (red.), *Work hours and work addiction* (pp. 193-252). Northampton (UK): Edward Elgar.
- Schaufeli, W.B., Taris, T.W., & Bakker, A.B. (2008a). It takes two to tango: Workaholism is working excessively and working compulsively. In R.J. Burke & C.L. Cooper (red.), *The long work hours culture: Causes, consequences and choices* (pp. 203-226). Bingley, (UK): Emerald.
- Schaufeli, W.B., Taris, T.W., & Van Rhenen, W. (2008b). Workaholism, burnout and engagement: Three of a kind or three different kinds of employee well-being. *Applied Psychology: An International Review*, 57, 173-203.
- Schaufeli, W.B., Shimazu, A., & Taris, T.W. (2009). Being driven to work excessively hard. The evaluation of a two-factor measure of workaholism in the Netherlands and Japan. *Cross-Cultural Research*, 43, 320-348.
- Schaufeli, W.B., Leiter, M.P., Maslach, C., & Jackson, S.E. (1996). Maslach burnout inventory – general survey. In C. Maslach, S.E. Jackson & M.P. Leiter (red.), *The Maslach burnout inventory: Test manual* (3rd ed., p. 22-26). Palo Alto: Consulting Psychologists Press.
- Schaufeli, W.B., Salanova, M., González-Romá, V., & Bakker, A.B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies*, 3, 71-92.
- Scott, K.S., Moore, K.S., & Miceli, M.P. (1997). An exploration of the meaning and consequences of workaholism. *Human Relations*, 50, 287-314.
- Siegrist, J. (1998). Adverse health effects of effort-reward imbalance at work. In C.L. Cooper (red.), *Theories of organizational stress* (pp. 190-204). Oxford (UK): Oxford University Press.
- Snir, R. & Zohar, D. (2008). Workaholism as discretionary time investment at work: An experience-sampling study. *Applied Psychology: An International Review*, 57, 109-127.
- Spence, J.T. & Robbins, A.S. (1992). Workaholism: Definition, measurement, and preliminary results. *Journal of Personality Assessment*, 58, 160-178.

- Stoeber, J., Otto, K., & Dalbert, C. (2009). Perfectionism and the big five: Conscientiousness predicts longitudinal increases in self-oriented perfectionism. *Personality and Individual Differences*, 47, 363-368.
- Taris, T.W. & Schaufeli, W.B. (2007). Workaholisme. In W.B. Schaufeli & A.B. Bakker (red.), *De psychologie van arbeid en gezondheid* (pp. 359-372). Houten: Bohn Stafleu Van Loghum.
- Taris, T.W., Van Beek, I., & Schaufeli, W.B. (2010). Perfectionism, workaholism, and burnout: A mediational analysis. *Romanian Journal of Applied Psychology*.
- Taris, T.W., Schaufeli, W.B., & Shimazu, A. (2010). The push and pull of work: About the difference between workaholism and work engagement. In A.B. Bakker & M.P. Leiter (red.), *Work engagement: A handbook of essential theory and research* (pp. 39-53). New York: Routledge.
- Taris, T.W., Schaufeli, W.B., & Verhoeven, L.C. (2005). Internal and external validation of the Dutch Work Addiction Risk Test: Implications for jobs and non-work conflict. *Journal of Applied Psychology: An international Review*, 54, 37-60.
- Taris, T.W., Geurts, S.A.E., Schaufeli, W.B., Blonk, R.W.B., & Lagerveld, S. (2008). All day and all of the night: The relative contribution of two dimensions of workaholism to well-being in self-employed workers. *Work & Stress*, 22, 153-165.
- Unwin, T. (2004). Gustave Flaubert, the hermit of Croisset. In T. Unwin (red.), *The Cambridge companion to Flaubert* (pp. 1-24). Cambridge (UK): Cambridge University Press.
- Vallerand, R.J., Blanchard, C.M., Mageau, G.A., Koestner, R., Ratelle, C., Leonard, M., & Gagné, M. (2003). Les passions de l'âme: On obsessive and harmonious passion. *Journal of Personality and Social Psychology*, 85, 756-767.
- Van Beek, I., Taris, T.W., & Schaufeli, W.B. (2010). *Motivation and job related well-being: Why do workaholic, engaged and burned-out employees work so hard*. Manuscript aangeboden ter publicatie.
- Van Veldhoven, M., De Jonge, J., Broersen, S., Kompier, M., & Meijman, T. (2002). Specific relationships between psychosocial job conditions and job-related stress: A three-level analytic approach. *Work & Stress*, 16, 207-228.
- Van Wijhe, C., Peeters, M.C.W., & Schaufeli, W.B. (2010). Understanding and treating workaholism: Setting the stage for successful interventions. In: R.J. Burke & C. Cooper (red.), *Risky Business: Psychological, physical and financial costs of high risk behaviour in organizations* (pp. 107-134). Farnham: Ashgate.

Ervaren Uitdaging op het Werk

Ontwikkeling van een Meetinstrument

*Paul T. Y. Preenen, Annelies E. M. van Vianen, Irene E. de Pater & René Geerling**

In twee studies werd een meetinstrument van ervaren uitdaging ontwikkeld en psychometrisch onderzocht. Ervaren uitdaging bleek uit drie onderscheidbare factoren te bestaan: positieve stimulatie, competentietoetsing en onzekerheid. Deze subschalen van ervaren uitdaging waren betrouwbaar en vertoonden een goede convergente validiteit. Niet-significante correlaties tussen ervaren uitdaging en de Big Five persoonlijkheidseigenschappen vriendelijkheid en neuroticisme wezen op een goede discriminante validiteit. De criteriumvaliditeit van het meetinstrument was eveneens goed. Vooral de subschaal positieve stimulatie correleerde aanzienlijk met de werkattitudes van respondenten. Het instrument voor ervaren uitdaging lijkt bruikbaar voor onderzoek en praktijk.

1. Inleiding

Uitdaging op het werk is voor werknemers en organisaties een populair en belangrijk begrip. Zo staan personeelsadvertenties veelal vol met leuzen als: 'Ben jij op zoek naar een nieuwe uitdaging?' en 'Ga jij de uitdaging aan?' (zie o.a. Monsterboard.nl; Nationalevacaturebank.nl). Werknemers vinden uitdagend werk blijkbaar aantrekkelijk aangezien het een belangrijke rol speelt bij het kiezen van een baan (Slaughter, Richard, & Martin, 2006) en bij de keuze van blijven of vertrekken bij de organisatie (Preenen, De Pater & Van Vianen, 2008a, Preenen, De Pater, Van Vianen & Keijzer, 2011). Bovendien lijkt uitdaging op het werk voor werknemers en hun organisaties allerlei positieve gevolgen met zich mee te brengen. Zo blijkt uitdaging positief samen te hangen met leren en persoonlijke ontwikkeling (DeRue & Wellman, 2009; Dragoni, Tesluk, Russell, & Oh, 2009; McCauley, Ruderman, Ohlott, & Morrow, 1994), promotiemogelijkheden (De Pater, Van Vianen, Bechtoldt, & Klehe, 2009) en positieve werkattitudes van werknemers (Cavanaugh, Boswell, Roehling, & Boudreau, 2000; Podsakoff, Lepine, & Lepine, 2007).

* Dr. Paul T.Y. Preenen was ten tijde van dit onderzoek verbonden aan de Programmagroep Arbeids- en Organisationspsychologie, Universiteit van Amsterdam. Prof. Dr. Annelies E.M. Van Vianen is verbonden aan de Organisationspsychologie, Universiteit van Amsterdam. Dr. Ing. Irene E. De Pater is verbonden aan Programmagroep Arbeids- en Organisationspsychologie, Universiteit van Amsterdam. René Geerling was ten tijde van dit onderzoek student-assistent bij de Programmagroep Arbeids- en Organisationspsychologie, Universiteit van Amsterdam.

Het begrip uitdaging wordt in de literatuur op verschillende manieren geoperationaliseerd (Preenen, De Pater, & Van Vianen, 2008b): als 'objectief' kenmerk van een taak of als een subjectieve ervaring van uitdaging. Het meeste onderzoek operationaliseert uitdaging in termen van 'objectieve' taakkenmerken zoals het uitvoeren van taken (1) die nieuw zijn, (2) die vaardigheden vereisen die een persoon nog niet volledig beheerst, (3) waarbij iemand veel autonomie heeft, en (4) die vaak een grotere verantwoordelijkheid en zichtbaarheid met zich meebrengen (zie Van Vianen, De Pater, & Preenen, 2008). Voor het meten van uitdagende, 'objectieve' werkkenmerken zijn er een aantal goede valide en betrouwbare vragenlijsten ontwikkeld, zoals de Developmental Challenge Profile (DCP; McCauley et al., 1994; McCauley, Ohlott, & Ruderman, 1999) of de Job Diagnostic Survey (Hackman & Oldham, 1980).

Sommige onderzoekers hebben uitdaging voornamelijk geoperationaliseerd in termen van beleving van medewerkers, dus als ervaren uitdaging (De Jonge, Dollard, Dormann, Le-Blanc, & Houtman, 2000; Hall & Lawler, 1970; Holmes & Srivastava, 2002; Idsoe, 2006; Walsh, Taber, & Beehr, 1980). Dit laatste is begrijpelijk aangezien subjectieve waarnemingen nog meer dan objectieve feiten van invloed zijn op de werkattitudes van mensen (Judge, Bono, & Locke, 2000). Ervaren uitdaging zou voor sommige uitkomsten dus wel eens een belangrijkere voor-speller kunnen zijn dan objectief uitdagende kenmerken.

Interessant genoeg ontbreekt het tot op heden aan een gevalideerd en empirisch onderbouwd meetinstrument voor de subjectieve ervaring van uitdaging. Daarnaast zijn de bestaande, ongevalideerde schalen die onderzoekers tot nog toe gebruikten om ervaren uitdaging te meten te eenzijdig. Uit de literatuur komt naar voren dat uitdaging geassocieerd wordt met positieve (e.g., Csikszentmihalyi, 1997; Lazarus & Folkman, 1984; Meyer & Allen, 1988) en negatieve (McCauley et al., 1994; Taylor, 1981; Van Vianen et al., 2008) emoties en ervaringen, alsmede met het idee dat capaciteiten getest en gebruikt worden (e.g., Hackman & Oldham, 1976; Jones & James, 1979). Kortom, ervaren uitdaging is een concept dat verschillende kernaspecten bevat.

Echter, tot nog toe zijn de meeste schalen enkel gericht op het leren, testen en gebruiken van competenties. De vragenlijst van Walsh et al. (1980) richt zich bijvoorbeeld op percepties van leermogelijkheden ('Op mijn werk krijg ik een kans mijn vaardigheden te gebruiken' en 'Mijn baan vereist dat ik nieuwe dingen blijf leren'). Ook Hall en Lawler (1970) gebruikten een vragenlijst gericht op het meten van leermogelijkheden ('Mijn baan geeft me de kans nieuwe vaardigheden en technieken te leren'), en de mate waarin de baan passend wordt gevonden ('Gegeven mijn opleiding, heb ik een baan die bij mijn mogelijkheden past'). Ettington (1998) daarentegen richt zich juist bovenal op de gevoelens van ervaren uitdaging.

Daarnaast operationaliseerden sommige onderzoekers ervaren uitdaging slechts aan de hand van enkele items (e.g., Taylor, 1981) of zelfs maar één item (e.g., De Jonge et al., 2000). Een belangrijke tekortkoming van single-item measures is dat de betrouwbaarheid niet getoetst kan worden (Wanous, Reichers, & Hudy, 1997). Daarnaast rijst de vraag of je een complex construct als ervaren uitdaging kunt vatten met één item dan wel enkele items.

Tett en collega's bepleiten juist het ontwikkelen van meerdere specifieke subschalen van complexe constructen (Tett, Guterman, Bleier, & Murphy, 2000).

Gezien het belang van ervaren uitdaging voor individuele medewerkers en hun organisaties, het feit dat een betrouwbaar, valide instrument van ervaren uitdaging tot op heden ontbreekt, en dat bestaande maten deels tekortschieten, hebben we een meetinstrument van ervaren uitdaging ontwikkeld en psychometrisch onderzocht. Omdat bestaande maten voor ervaren uitdaging te eenzijdig, dan wel te generiek waren, zullen wij een vragenlijst ontwikkelen waarin de verschillende aspecten van ervaren uitdaging aan bod komen. Zodoende creëren we een instrument dat kan helpen om een genuanceerder beeld van het praktisch erg relevante concept ervaren uitdaging, alsmede de specifieke gevolgen van ervaren uitdaging te verkrijgen.

In een eerste onderzoek (Studie 1) genereerden we de items en onderzochten we de factorstructuur van de items. In een tweede onderzoek (Studie 2) onderzochten we de factoriële validiteit, interne betrouwbaarheid, test-heretest betrouwbaarheid en de constructvaliditeit van het instrument voor het meten van ervaren uitdaging. Hierbij hebben we ons laten leiden door de procedures voor instrumentvalidatie zoals deze in de psychometrische literatuur worden beschreven (e.g., Cortina, 1993; Robinson, Shaver, & Wrightsman, 1991; Schriesheim, Powers, Scandura, Gardiner, & Lankau, 1993; Schwab, 1999).

2. Studie 1

In Studie 1 genereerden we de items en onderzochten de factorstructuur en interne consistentie van een eerste versie van het meetinstrument voor ervaren uitdaging.

Om een zo goed en uitputtend mogelijke itemset te genereren werd de eerste selectie van items ontleend aan eerder onderzoek (Preenen et al., 2008b), waarvan het doel was om een wetenschappelijk gefundeerde, uitputtende en bottom-up conceptualisatie van werkuitdaging te ontwikkelen. Hierin werd aan 132 respondenten gevraagd om een recent uitgevoerde uitdagende taak te beschrijven en uit te leggen waarom zij de taak uitdagend vonden. Dit resulteerde in 233 unieke beschrijvingen van uitdagende taken. Vervolgens werden deze beschrijvingen met een concept mapping techniek geanalyseerd (e.g., Behfar, Peterson, Mannix, & Trochim, 2008; Trochim, 1989). Dit hield in dat eerst tien onafhankelijke personen de beschrijvingen sorteerden in overeenkomstige beschrijvingen. De resultaten werden verwerkt in een binary square matrix, waarop vervolgens een multiple dimensional scaling analysis en een hiërarchische clusteranalyse werden uitgevoerd. Dit resulteerde in 37 deelaspecten van werkuitdaging die vervolgens door drie expertonderzoekers op basis van theoretische overwegingen en inhoudelijke discussie werden gecategoriseerd als: (1) taakkenmerken, (2) omgevingkenmerken (3) de cognitieve beleving van uitdaging en (4) de affectieve beleving van uitdaging (emoties en gevoelens bij uitdaging). De eerste twee categorieën zijn op te vatten als 'objectieve' uitdagende taakkenmerken die de subjectieve staat van uitdaging, categorie 3 en 4, beïnvloeden. De beschrijvingen van de

derde en vierde categorie verwijzen naar ervaren uitdaging, welke we als uitgangspunt hebben gebruikt voor de items van ons instrument. Samen bevatten de twee categorieën 16 deelaspecten met 113 unieke beschrijvingen van uitdagende ervaringen.

Deze beschrijvingen sluiten goed aan bij de bestaande uitdagingliteratuur, waarin uitdaging geassocieerd wordt met positieve (e.g., Csikszentmihalyi, 1997; Lazarus & Folkman, 1984; Meyer & Allen, 1988) en negatieve (McCauley et al., 1994; Taylor, 1981; Van Vianen et al., 2008) emoties en ervaringen, alsmede met het idee dat capaciteiten getest en gebruikt worden (e.g., Hackman & Oldham, 1976; Jones & James, 1979).

Om het grote aantal beschrijvingen te reduceren werden beschrijvingen die een sterke overlap vertoonden door ons (drie expertonderzoekers) samengevoegd. We vatten bijvoorbeeld de beschrijvingen 'Ik wil aan mezelf bewijzen dat ik het kan doen' en 'Jezelf moeten bewijzen' samen tot het volgende item: 'In mijn werk voer ik taken uit waarbij ik mezelf moet bewijzen', en we vatten begrippen zoals 'zwaar', 'hoge lat' en 'moeilijk' samen tot: 'In mijn werk voer ik taken uit die moeilijk haalbaar zijn.' In een tweede stap verwijderden we beschrijvingen die een lage inhoudsvaliditeit hadden en die niet aansloten bij de literatuur. We verwijderden bijvoorbeeld de beschrijving 'Iets doen wat ik altijd al wilde doen'.

De reductie mondde uit in een uiteindelijke set van 19 unieke en bruikbare beschrijvingen die in items werden omgezet. Veertien items hebben betrekking op de cognitieve beleving van uitdaging en vijf items verwijzen naar de affectieve beleving¹.

Om te onderzoeken of de items in verschillende dimensies zijn op te delen werd eerst de factorstructuur onderzocht. Vervolgens werd de interne consistentie van het meetinstrument onderzocht.

2.1 *Methode Studie 1*

Deelnemers en Procedure

Vijfhonderd deelnemers aan een internetpanelstudie werden per e-mail verzocht deel te nemen aan het onderzoek. Tweehonderd tweeëntwintig (123 mannen, 99 vrouwen) deelnemers (44%) vulden de vragenlijst in en verstrekten demografische gegevens. De gemiddelde leeftijd van de respondenten was 42,05 jaar (SD = 8.53) en hun gemiddelde dienstverband was 5,42 jaar (SD = 5.87). De meeste respondenten (N = 182) hadden een hbo- of universitair diploma en 40 deelnemers hadden een mbo- of lagere opleiding afgerond. De respondenten werkten gemiddeld 34,34 uur per week (SD = 5.73).

De items werden gemeten met een 7-punts Likert-schaal, variërend van 1 (volledig mee oneens) tot 7 (volledig mee eens).

2.2 *Resultaten Studie 1*

Om de onderliggende factorstructuur van de items vast te stellen werd een exploratieve factoranalyse met varimax-rotatie (principale componenten methode) uitgevoerd. De analyse resulteerde in drie factoren met eigenwaarden > 1, die samen 67,78% van de variantie verklaarden. Twee items ('In mijn werk voer ik taken uit waarin ik nieuwe ervaringen en impressies krijg' en 'In mijn werk voer ik taken uit

Tabel 1 Exploratieve Factoranalyse van Ervaren Uitdaging (Studie 1, N = 222)

Item	Factor 1 Positieve stimulatie	Factor 2 Competentie- toetsing	Factor 3 Onzekerheid
In mijn werk voer ik taken uit:			
1. die ik aantrekkelijk vind	.86	.06	-.06
2. die ik interessant vind	.86	.10	.03
3. die ik uitdagend vind	.80	.18	.33
4. die me stimuleren	.78	.19	.28
5. die ik zinvol vind	.85	.09	-.02
6. die ik boeiend vind	.90	.09	.05
7. waarbij ik mezelf moet bewijzen	.22	.74	.21
8. waarbij ik mijzelf moet overwinnen	-.08	.67	.29
9. waarbij mijn kunde wordt getest	.28	.83	.06
10. waarbij ik op de proef word gesteld	.08	.89	.15
11. waarbij ik een stapje verder moet gaan dan normaal	.08	.77	.33
12. waarbij ik veel van mijzelf moet geven	.22	.67	.36
13. waarbij het niet altijd zeker is of ik het kan	-.03	.17	.78
14. waarbij ik het risico loop te falen	-.17	.35	.68
15. die moeilijk haalbaar zijn	.42	.31	.59
16. die lastig haalbaar zijn	.13	.31	.62
17. waarbij ik te maken heb met taken die nieuw voor mij zijn	.35	.13	.73
Eigenwaarden	6.92	3.37	1.39
% verklaarde variantie	40.69	19.84	8.20

waarbij ik obstakels moet overkomen') laadden hoog op twee factoren en werden derhalve verwijderd. Een aanvullende exploratieve factoranalyse over de resterende 17 items resulteerde in drie duidelijk interpreteerbare factoren die 68,73% van de variantie konden verklaren. De factorladingen, eigenwaarden en verklaarde varianties zijn weergegeven in Tabel 1. De eerste factor bestond uit zes items die taakwaardering en ervaren stimulatie uitdrukken. De factorladingen van de items varieerden van .78 tot .90. We noemen deze factor de subschaal 'positieve stimulatie' van ervaren uitdaging. De tweede factor bestond uit zes items die verwijzen naar de beleving getest te worden en zichzelf te moeten bewijzen. De factorladingen van deze items varieerden van .67 tot .89. We noemen deze subschaal 'competentietoetsing'. De derde factor bestond uit vijf items die te maken hebben met het ervaren van risico, moeilijkheid en onzekerheid. De factorladingen van deze items varieerden van .59 tot .78. Deze subschaal duiden we aan met 'onzekerheid'.

Tabel 2 Beschrijvende Statistieken en Correlaties tussen de Subscales van Ervaren Uitdaging (Studie 1, N = 222)^a

	M	SD	1	2	3
1. Positieve stimulatie	5.26	.85	(.93)		
2. Competentietoetsing	4.48	.98	.33	(.89)	
3. Onzekerheid	4.17	.93	.33	.60	(.80)

Noot. a Alle correlaties zijn significant ($p < .01$). Cronbach's alfa's staan tussen haakjes weergegeven op de diagonaal.

De gemiddelden, standaardafwijkingen, correlaties en interne consistenties van de drie subscales zijn weergegeven in tabel 2. Omdat de drie subscales tot het bredere construct van ervaren uitdaging behoren is te verwachten dat ze een onderlinge samenhang zullen vertonen. Dit bleek het geval te zijn. Positieve stimulatie correleerde redelijk hoog met zowel competentietoetsing als onzekerheid (.33, $p < .01$). Een hogere correlatie werd echter gevonden tussen competentietoetsing en onzekerheid (.60, $p < .01$). De interne consistentie van de drie waargenomen subscales varieerde van .80 tot .93.

3. Studie 2

Op grond van op de resultaten van Studie 1 voerden we een tweede studie uit om de psychometrische kwaliteit van het meetinstrument verder te onderzoeken. Een goed meetinstrument moet voldoen aan de voorwaarden van betrouwbaarheid en validiteit. Betrouwbaarheid heeft betrekking op de nauwkeurigheid van het instrument en geeft een indicatie van de mate waarin de scores vrij zijn van meetfouten (Schwab, 1999). Voor het beoordelen van de betrouwbaarheid van het meetinstrument voor ervaren uitdaging zal de interne consistentie en test-hertest betrouwbaarheid van het instrument worden onderzocht. Interne consistentie betreft de gelijkheid van itemscores binnen een subschaal, terwijl de test-hertest betrouwbaarheid betrekking heeft op de consistentie van de meetresultaten bij herhaalde afname van het instrument (Schwab, 1999). In Studie 2 werd de test-hertest betrouwbaarheid gemeten met een interval van zes maanden.

De validiteit van een instrument geeft aan of de test meet wat hij beoogt te meten (Schwab, 1999). De validiteit wordt bepaald aan de hand van de factoriële validiteit en het nomologisch netwerk. De factoriële validiteit betreft de dimensionaliteit van het construct (Evers, 2001). Als het meetinstrument een betrouwbare en valide meting geeft van het te meten construct, moet de factorstructuur overeenkomen met de voorspellingen. Op basis van de resultaten van Studie 1 veronderstelden we dat ervaren uitdaging een hogere orde construct is dat drie subscales omvat.

Het nomologisch netwerk van ervaren uitdaging specificeert de relaties tussen ervaren uitdaging en andere constructen (Schwab, 1999). Voor het onderzoeken van het nomologisch netwerk wordt een onderscheid gemaakt in convergente validiteit, discriminante validiteit en criteriumvaliditeit. Convergente validiteit

heeft betrekking op relaties tussen verschillende metingen van hetzelfde construct. Er is sprake van convergente validiteit wanneer het meetinstrument sterk samenhangt met andere metingen van hetzelfde construct (Schwab, 1999). Dit impliceert dat er sterke relaties moeten bestaan tussen de subschalen van het meetinstrument voor ervaren uitdaging en verwante instrumenten, zoals de baanuitdaging-vragenlijst van Ettington (1998) en de uitdagende opdrachten-items van Ragins en McFarlin (1990). Daarnaast zou ervaren uitdaging moeten samenhangen met ervaren autonomie en ervaren variëteit in het werk (Evans & Kersh, 2004; Hackman & Oldham, 1980).

Discriminante validiteit heeft betrekking op de mate waarin ervaren uitdaging zich onderscheidt van andere constructen (Schwab, 1999). De discriminante validiteit van het meetinstrument voor ervaren uitdaging zal worden bepaald door de drie subschalen te correleren met twee Big Five persoonlijkheidskenmerken (Costa & McCrae, 1992), namelijk neuroticisme (angstig, defensief, onzeker en emotioneel) en vriendelijkheid (altruïstisch, warm, genereus en coöperatief). Zowel op theoretische als empirische gronden is er geen directe reden aan te nemen dat er sterke relaties bestaan tussen ervaren uitdaging en deze persoonlijkheidskenmerken. Enerzijds is het mogelijk dat neurotische individuen juist trachten uitdagende situaties te vermijden en daarom minder uitdaging ervaren op het werk, anderzijds zullen zij soms eerder dan anderen uitdaging ervaren omdat ze zich onzeker voelen. Dit maakt het dus lastig om verbanden te vinden. Derhalve verwachten wij geen relatie tussen neuroticisme en ervaren uitdaging. Vriendelijke mensen zouden misschien meer uitdagende taken uitvoeren omdat ze aardig zijn en hun dergelijke taken gegund worden door anderen en daardoor meer uitdaging ervaren. Anderzijds kan het ook zo zijn dat zij juist eerder genoegen nemen met minder uitdagende taken en de uitdagende taken gunnen aan anderen die graag zulke taken uitvoeren. Wij verwachten dus geen duidelijke relatie te vinden tussen vriendelijkheid en ervaren uitdaging.

Bovenstaande verwachtingen worden indirect ondersteund door empirisch onderzoek naar de relatie tussen de Big Five persoonlijkheidskenmerken en een aan uitdaging verwante variabele, namelijk motivatie tot leren (Major, Turner, & Fletcher, 2006). Uit dit onderzoek bleek dat vriendelijkheid en neuroticisme niet gerelateerd waren aan motivatie tot leren.

Er is sprake van criteriumvaliditeit wanneer ervaren uitdaging samenhangt met bepaalde uitkomstvariabelen en deze andere variabelen bij voorkeur kan voorspellen. Als de uitkomstvariabele gelijktijdig met de criteriumvariabele wordt gemeten spreekt men ook wel van concurrente validiteit (Cronbach & Meehl, 1955). Verscheidene onderzoeken hebben aangetoond dat uitdaging samenhangt met werkattitudes zoals werktevredenheid (e.g., Judge et al., 2000; Kirk-Brown & Wallace, 2004), affectieve commitment (e.g., Allen & Meyer, 1990; Dixon, Cunningham, Sagas, Turner, & Kent, 2005) en vertrekintenties (Podsakoff et al., 2007; Preenen et al., 2008a). We verwachtten daarom dat ervaren uitdaging positief met werktevredenheid en commitment, en negatief met vertrekintentie zou samenhangen.

Uit onderzoek van De Pater et al. (2009) bleek dat het uitvoeren van uitdagende taken niet samenhangt met de beoordeelde werkprestaties van medewerkers. In

dit onderzoek werd uitdaging echter geoperationaliseerd in termen van meer objectieve werkuitdagingaspecten. In het huidige onderzoek zullen we het verband tussen de subschalen van ervaren uitdaging en beoordeelde werkprestatie nader onderzoeken om te exploreren of de uitdagingdimensies wellicht wel samenhangen met beoordeelde werkprestatie. Dit zou de predictieve validiteit van de ervaren werkuitdaginglijst ten goede komen. Enerzijds zouden de subschalen competentietoetsing en onzekerheid mogelijk negatief kunnen samenhangen met werkprestatie. Immers, als werknemers een kloof ervaren tussen de taak-eisen en hun capaciteiten, en dus onzeker zijn over het volbrengen van de taak, kan dit een negatief effect hebben op de prestatiebeoordeling door hun leidinggevende. Anderzijds zou het kunnen zijn dat mensen die veel uitdaging ervaren en dus wellicht veel uitdagende, belangrijke taken op zich nemen en krijgen, hierdoor juist positief worden beoordeeld door hun leidinggevende. Het is evenzeer denkbaar dat juist de subschaal positieve stimulatie positief samenhangt met de prestatiebeoordeling, aangezien enthousiasme en inzet van de medewerker wellicht leiden tot een gunstige beoordeling door de leidinggevende. Om na te gaan hoe de verschillende schalen samenhangen met werkprestatie zullen wij ter exploratie de relatie tussen de ervaren uitdagingschalen en de door de leidinggevende beoordeelde werkprestatie onderzoeken.

3.1 *Methode Studie 2*

- *Deelnemers en Procedure*

De data voor dit onderzoek werden op twee tijdstippen verzameld. Op het eerste tijdstip werden 773 werknemers van verschillende Nederlandse bedrijven uit de commerciële dienstensector per e-mail uitgenodigd een online vragenlijst in te vullen. In totaal deden 468 mensen mee (61%), van wie 225 deelnemers man (48%) en 243 vrouw (52%) waren. De gemiddelde leeftijd was 30,38 jaar ($SD = 10.65$) en het gemiddelde dienstverband was 4,30 jaar ($SD = 6.44$). Van hen hadden 361 werknemers een hbo- of universitair diploma (77%) en 107 deelnemers een mbo- of lagere opleiding (23%). De respondenten werkten gemiddeld 28,19 uur per week ($SD = 12.68$).

Aan het onderzoek deden 16 leidinggevendenden mee die de werktevredenheid, commitment en prestaties van enkele van hun medewerkers ($N = 48$) beoordeelden. Van deze 48 beoordeelde medewerkers waren 32 man (67%) en 16 vrouw (33%). Hun gemiddelde leeftijd was 35,06 jaar ($SD = 10.6$) en hun gemiddelde dienstverband was 3,53 jaar ($SD = 3.25$). Van deze medewerkers hadden er 24 een hbo- of universitair diploma (50%) en 24 deelnemers hadden een mbo- of lagere opleiding (50%). Ze werkten gemiddeld 35,34 uur per week ($SD = 9.67$).

Op het tweede tijdstip, zes maanden later, werden per e-mail 119 follow-up vragenlijsten gestuurd naar respondenten die op het eerste tijdstip aangegeven hadden geïnteresseerd te zijn in deelname aan een follow-up studie. We vroegen respondenten nogmaals naar hun ervaren uitdaging. In totaal deden 59 deelnemers mee (50%). Van de deelnemers waren 28 man (48%) en 31 vrouw (52%). Hun gemiddelde leeftijd was 35,96 jaar ($SD = 11.96$) en hun gemiddelde dienstverband was 6,14 jaar ($SD = 7.58$). Deelnemers werkten gemiddeld 31,21 uur per week

(SD = 10.33). Van hen hadden 52 een hbo- of universitair diploma (88%) en 7 een mbo- of lagere opleiding (12%).

• *Meetinstrumenten*

Ervaren uitdaging. De subschalen positieve stimulatie, competentietoetsing en onzekerheid werden gemeten met dezelfde 17 items als in Studie 1 (zie tabel 1). Bij het beantwoorden van de items werd een 7-puntsschaal gebruikt, van 1 (volledig mee oneens) tot 7 (volledig mee eens). Cronbach's alfa voor positieve stimulatie was .95 (tijdstip 1) en .95 (tijdstip 2), voor competentietoetsing .89 (tijdstip 1) en .90 (tijdstip 2) en voor onzekerheid .86 (tijdstip 1) en .89 (tijdstip 2).

Baanuitdaging werd gemeten met de schaal van Ettington (1998), die deze items ontleende aan een schaal van Smith, Kendall en Hulin (1969). Aan respondenten werd gevraagd hoe vaak ze de volgende vijf woorden/termen zouden gebruiken om hun werk te beschrijven: (1) fascinerend, (2) routinematig, (3) saai, (4) creatief en (5) uitdagend. De antwoordopties varieerden van 1 (nooit) tot 5 (altijd). De antwoorden op de items 2 en 3 werden omgescoord, zodat hogere scores een hogere mate van baanuitdaging aangaven. Cronbach's alfa was .83.

Uitdagende opdrachten werd gemeten met de uit drie items bestaande subschaal 'uitdagende opdrachten' van het Mentor Role Instrument (MRI), ontwikkeld door Ragins en McFarlin (1990). Aan respondenten werd gevraagd aan te geven in hoeverre hun leidinggevende hun taken toewees zoals: (1) uitdagende opdrachten, (2) taken die het ontwikkelen van nieuwe vaardigheden vereisen en (3) taken die het leren van nieuwe vaardigheden met zich meebrengen. De antwoordopties varieerden van 1 (nooit) tot 5 (altijd). Cronbach's alfa was .93.

Autonomie werd gemeten met een subschaal van de Job Diagnostic Survey (Hackman & Oldham, 1975). Een voorbeelditem is: 'Mijn werk geeft me aanzienlijke vrijheid en autonomie in hoe ik mijn werk uitvoer.' De antwoordopties varieerden van 1 (volledig mee oneens) tot 5 (volledig mee eens). Cronbach's alfa was .73.

Taakvariatie werd gemeten met een subschaal van de Job Diagnostic Survey (Hackman & Oldham, 1975). Een voorbeelditem is: 'Mijn werk vereist dat ik een aantal vaardigheden gebruik die complex of van een hoog niveau zijn.' De antwoordopties varieerden van 1 (volledig mee oneens) tot 5 (volledig mee eens). Cronbach's alfa was .74.

Neuroticisme werd gemeten met de uit zes items bestaande schaal 'neuroticisme' van de Big Five 'adjective marker' (Sheldon, Ryan, Rawsthorne, & Ilardi, 1997). Respondenten gaven aan in hoeverre de items (1) ongelukkig, (2) onzeker, (3) vol zelfvertrouwen (omgescoord), (4) vrolijk (omgescoord), (5) somber en (6) humeurig op hen van toepassing zijn, op een schaal van 1 (helemaal niet op mij van toepassing) tot 5 (helemaal op mij van toepassing). Cronbach's alfa was .65.

Vriendelijkheid werd gemeten met de uit zes items bestaande vriendelijkheids-schaal van de Big Five 'adjective marker' (Sheldon et al., 1997). Respondenten gaven aan hoe accuraat de items (1) attent, (2) aardig, (3) vriendelijk, (4) coöperatief, (5) geduldig en (6) egocentrisch (omgescoord) henzelf beschrijven, op een schaal van 1 (helemaal niet op mij van toepassing) tot 5 (helemaal op mij van toepassing). Cronbach's alfa was .65.

Werktevrede werd gemeten met vijf items (Brayfield & Rothe, 1951). Respondenten gaven op een 5-puntsschaal aan in hoeverre ze het eens zijn met de items, variërend van 1 (volledig mee oneens) tot 5 (volledig mee eens). Een voorbeelditem van de medewerkersvragenlijst ($\alpha = .86$) is: 'Ik voel me tevreden met mijn baan.' Een voorbeelditem van de leidinggevendenvragenlijst ($\alpha = .79$) is: 'Deze medewerker voelt zich tevreden met zijn/haar baan.'

Commitment werd gemeten met vijf items die affectieve commitment meten (De Gilder, Van den Heuvel, & Ellemers, 1997). Respondenten gaven op een 5-puntsschaal aan in hoeverre ze het eens zijn met de items, variërend van 1 (volledig mee oneens) tot 5 (volledig mee eens). Een voorbeelditem van de medewerkersvragenlijst ($\alpha = .90$) is: 'Deze organisatie betekent veel voor me.' Een voorbeelditem van de leidinggevendenvragenlijst ($\alpha = .94$) is: 'Deze organisatie betekent veel voor deze medewerker.'

Vertrekintentie werd gemeten met vijf items (Wayne, Shore, & Liden, 1997). Een voorbeelditem is: 'Ik denk er serieus over ontslag te nemen.' Deelnemers antwoordden op een 7-puntsschaal, variërend van 1 (volledig mee oneens) tot 7 (volledig mee eens). Cronbach's alfa was .85.

Werkprestatie werd gemeten met vijf items (Denison, Hooijberg, & Quinn, 1995). Leidinggevenden gaven op een 7-puntsschaal aan: (1) in hoeverre hun medewerkers net zo succesvol zijn als hun collega-medewerkers, (2) hun prestatienormen halen, (3) een rolmodel voor anderen zijn, (4) succesvol zijn en (5) in het algemeen effectief zijn. Cronbach's alfa was .93.

4. Resultaten Studie 2

4.1 *Confirmatieve Factoranalyse*

We voerden verschillende confirmatieve factoranalyses uit (LISREL 8.80; Jöreskog & Sörbom, 2006) om het 3-factormodel van ervaren uitdaging te vergelijken met alternatieve modellen: drie modellen met twee factoren en een model waarbij alle items op één factor laadden. We gebruikten verschillende indicatoren om vast te stellen in hoeverre een model bij de geobserveerde data past (de fit van het model), zoals de chi-kwadraat test (χ^2), de 'root mean square error of approximation' (RMSEA), de 'non-normed fit index' (NNFI), de 'comparative fit index' (CFI), de 'incremental fit index' (IFI) en de 'standardized root mean square residual' (SRMR). Er is sprake van een aanvaardbare fit wanneer de RMSEA en SRMR < .10 zijn en de NNFI, CFI en IFI > .95. Tabel 3 toont aan dat het 3-factormodel² waarbij correlaties tussen de drie latente variabelen werden geschat significant beter fit met de data dan de alternatieve modellen (zie $\Delta\chi^2$ in tabel 3). De factorladingen van positieve stimulatie varieerden van .76 tot .90, de factorladingen van competentietoetsing varieerden van .70 tot .82 en de factorladingen van onzekerheid varieerden van .65 tot .85. De data vertoonden aldus ondersteuning voor het onderscheiden van de drie subschalen van ervaren werkuitdaging.

Tabel 3 *Fit Statistieken van Confirmatorische Factoranalyse (Studie 2, N = 468)*

	χ^2	df	RMSEA	NNFI	CFI	IFI	SRMR	$\Delta \chi^2$ ^a	Δdf ^a	p ^a
1 3-factor model	503.87	113	.09	.97	.98	.98	.06			
2 2-factor model ^b	1990.75	115	.19	.92	.93	.93	.10	1486.88	2	.001
3 2-factor model ^c	1293.88	115	.15	.94	.95	.95	.11	790.01	2	.001
4 2-factor model ^d	676.66	115	.10	.96	.97	.97	.07	172.79	2	.001
5 1-factor model ^e	2111.11	116	.19	.90	.92	.92	.10	1607.24	3	.001

Noot. a Modellen 2 tot 5 zijn vergeleken met model 1.

Noot. b De factor onzekerheid en een factor die positieve stimulatie en competentietoetsing combineert.

Noot. c De factor competentietoetsing en een factor die positieve stimulatie en onzekerheid combineert.

Noot. d De factor positieve stimulatie en een factor die onzekerheid en competentietoetsing combineert.

Noot. e Een factor die onzekerheid, positieve stimulatie en competentietoetsing combineert.

Tabel 4 *Beschrijvende Statistieken en Correlaties tussen de Subscales van Ervaren Uitdaging (Studie 2)^a*

	Tijdstip	M	SD	1	2	3	4	5	6
1. Positieve stimulatie	1	5.09	1.30	(.95)					
2. Competentietoetsing	1	4.68	1.26	.57	(.89)				
3. Onzekerheid	1	4.16	1.30	.54	.72	(.86)			
4. Positieve stimulatie	2	5.05	1.31	.65	.45	.40	(.95)		
5. Competentietoetsing	2	4.64	1.30	.31	.46	.38	.53	(.90)	
6. Onzekerheid	2	4.21	1.32	.41	.46	.58	.49	.74	(.89)

Noot. a Tijdstip 1, N = 468; Tijdstip 2, N = 59. Alle correlaties zijn significant ($r = .31$, $p < .05$; $r > .31$, $p < .01$). Cronbach's alfa's staan tussen haakjes weergegeven op de diagonaal.

4.2 Betrouwbaarheid

Een goed instrument moet aan de eisen van interne consistentie en test-hertest betrouwbaarheid voldoen. Tabel 4 laat de gemiddelden, standaardafwijkingen, betrouwbaarheidscoëfficiënten en correlaties van de drie subscales van ervaren uitdaging zien, zoals gemeten op tijdstip 1 en tijdstip 2. In Studie 2 waren de Cronbach's alfa's van de drie subscales goed (variërend van .86 tot .95) en gelijk aan de waarden uit Studie 1. Deze resultaten laten zien dat de drie subscales van ervaren uitdaging intern consistent zijn. De test-hertest betrouwbaarheid werd bepaald aan de hand van de correlaties tussen de metingen van tijdstip 1 en 2. De test-hertest correlaties waren .65 voor positieve stimulatie, .46 voor competentietoetsing en .58 voor onzekerheid. Deze resultaten suggereren dat de subscales van ervaren uitdaging een matig tot redelijke test-hertest betrouwbaarheid hebben.

Tabel 5 *Correlaties tussen de Subscales van Ervaren Uitdaging en Studie 2 Variabelen*

	Positieve stimulatie	Competentie toetsing	Onzekerheid
Convergente validiteit			
Baanuitdaging	.80**	.53**	.51**
Uitdagende opdrachten	.52**	.46**	.49**
Autonomie	.44**	.22**	.24**
Taakvariatie	.69**	.49**	.52**
Discriminante validiteit			
Neuroticisme	-.08	.05	.09*
Vriendelijkheid	.04	-.00	-.04
Criteriumvaliditeit			
Werktevredenheid (zelfrapportage)	.68**	.26**	.22**
Commitment (zelfrapportage)	.43**	.23**	.23**
Vertrekintenties (zelfrapportage)	-.53**	-.25**	-.25**
Werktevredenheid (leidinggevende) ^a	.25†	.02	.10
Commitment (leidinggevende) ^a	.41**	.07	.04
Werkprestatie (leidinggevende) ^a	.04	-.27†	-.16

Noot. N = 468; ^a N = 48.

** p < .01, * p < .05, † p < .10

4.3 Validiteit

De convergente validiteit van het meetinstrument voor ervaren uitdaging werd onderzocht door de drie subscales te correleren met de items van baanuitdaging (Ettington, 1998), uitdagende opdrachten (Ragins & McFarlin, 1990), autonomie en taakvariatie. Tabel 5 laat zien dat de drie subscales van ervaren uitdaging en vooral de schaal 'positieve stimulatie', substantieel met de andere schalen van uitdaging correleren (correlaties met positieve stimulatie varieerden van .44 tot .80, p 's < .01). Deze resultaten bieden ondersteuning voor de convergente validiteit van het meetinstrument van ervaren uitdaging. De discriminante validiteit van het meetinstrument voor ervaren uitdaging werd onderzocht door de drie subscales te correleren met neuroticisme en vriendelijkheid. De correlaties waren nihil tot zeer laag (zie Tabel 5). De subscales van ervaren uitdaging lijken aldus een goede discriminante validiteit te hebben.

Om de criteriumvaliditeit (concurrente validiteit) van het meetinstrument te onderzoeken correleerden we de drie subscales van ervaren uitdaging met de werkattitudes van medewerkers. Tabel 5 laat zien dat de subscales van ervaren uitdaging aanzienlijk met werktevredenheid, commitment en vertrekintentie van werknemers correleerden (absolute r 's varieerden van .23 tot .68, alle p 's < .01). Vooral de correlaties met de schaal 'positieve stimulatie' waren hoog (absolute r 's varieerden van .43 tot .68). Deze resultaten geven een eerste indicatie van een goede criteriumvaliditeit van ervaren uitdaging. De werkattitudes van medewer-

kers waren echter gebaseerd op zelfrapportage waardoor er mogelijk sprake zou kunnen zijn van geflatteerde correlaties veroorzaakt door gedeelde methode variatie. Daarom onderzochten we tevens de relaties tussen de subschalen van ervaren uitdaging en het oordeel van leidinggevendenden over de werktevredenheid, commitment en werkprestatie van hun medewerkers (N = 48). De subschaal 'positieve stimulatie' had een substantiële correlatie met de door de leidinggevende beoordeelde werktevredenheid ($r = .25$, $p < .10$) en commitment ($r = .41$, $p < .01$) van medewerkers. Er werd geen samenhang gevonden met de beide andere subschalen van ervaren uitdaging. De subschaal 'competentietoetsing' was echter negatief gerelateerd aan de door de leidinggevende beoordeelde werkprestatie van medewerkers ($r = -.27$, $p < .10$). Met inachtneming van de kleine sample bieden deze resultaten enige aanvullende ondersteuning voor de criteriumvaliditeit van ervaren uitdaging.

5. Conclusie en discussie

Dit onderzoek had tot doel om een meetinstrument voor ervaren uitdaging te ontwikkelen en te valideren. In Studie 1 onderzochten we de factorstructuur van een eerste set items van ervaren uitdaging. In Studie 2 werden de factorstructuur, de betrouwbaarheid en de convergente, discriminante en criteriumvaliditeit onderzocht.

Op grond van de eerste studie werden drie factoren van ervaren uitdaging onderscheiden: (1) positieve stimulatie, dat refereert aan taakwaardering en ervaren stimulatie, (2) competentietoetsing, dat refereert aan de mate waarin mensen op de proef worden gesteld en zichzelf moeten bewijzen, en (3) onzekerheid, dat refereert aan de mate waarin mensen hun werktaken beschouwen als moeilijk en riskant, en onzeker zijn over het kunnen volbrengen ervan.

De resultaten van Studie 2 bevestigden de 3-factorstructuur van ervaren uitdaging. De drie subschalen correleren echter behoorlijk met elkaar. Opvallend is de relatief hoge correlatie tussen competentietoetsing en onzekerheid (r 's variërend van .60 tot .72), wat suggereert dat medewerkers die zich getest voelen ook onzeker zijn over hun mogelijkheid om de uitdagende taken te volbrengen.

De interne consistenties van de drie subschalen waren in beide studies goed, maar de test-hertest correlatie is voor verbetering vatbaar. De matige test-hertest betrouwbaarheid is mogelijk toe te schrijven aan de relatief lange periode tussen de beide metingen. In een periode van zes maanden kunnen werkervaringen aan verandering onderhevig zijn (De Pater et al., 2009; McDaniel, Schmidt, & Hunter, 1988). In toekomstig onderzoek zal binnen een kortere tijdsperiode de test-hertest betrouwbaarheid van ervaren uitdaging nader onderzocht moeten worden.

De resultaten van de tweede studie boden ondersteuning voor de convergente, discriminante en criteriumvaliditeit van het meetinstrument voor ervaren uitdaging. De subschalen correleerden substantieel met twee verwante schalen van uitdaging en met autonomie en taakvariatie, wat wijst op een goede convergente validiteit. De lage correlaties tussen de subschalen van ervaren uitdaging en de Big Five persoonlijkheidseigenschappen vriendelijkheid en neuroticisme suggere-

ren dat het meetinstrument voor ervaren uitdaging een goede discriminante validiteit heeft. De aanzienlijke correlaties tussen ervaren uitdaging en werkhoudingen bieden ondersteuning voor de criteriumvaliditeit (concurrerende validiteit) van het meetinstrument voor ervaren uitdaging. Bovendien werd er een positief verband gevonden tussen de subschaal 'positieve stimulatie' en door de leidinggevende beoordeelde werkhoudingen van medewerkers. Deze bevindingen komen overeen met eerder onderzoek dat uitdaging relateerde aan werktevredenheid (e.g., Judge et al., 2000; Kirk-Brown & Wallace, 2004), commitment (e.g., Allen & Meyer, 1990; Buchanan, 1974; Dixon et al., 2005), vertrekintenties en daadwerkelijk vertrek (e.g., Podsakoff et al., 2007; Preenen et al., 2008a; Preenen et al., 2011). We onderzochten ook de relatie tussen ervaren uitdaging en door de leidinggevende beoordeelde werkprestatie. We vonden een marginaal significante negatieve relatie tussen competentietoetsing en beoordeelde werkprestatie. Leidinggevers beoordeelden de werkprestatie lager als medewerkers rapporteerden dat ze in hun werk taken uitvoeren waarbij hun competenties worden getest. Deze medewerkers beschikten kennelijk nog niet over voldoende vaardigheden om hun taak optimaal uit te kunnen voeren. Opgemerkt moet worden dat deze bevinding op een kleine sample gebaseerd is en aldus gerepliceerd dient te worden in vervolgonderzoek.

5.1 Implicaties, tekortkomingen en toekomstig onderzoek

In de inleiding van dit artikel hebben we bestaande meetinstrumenten voor ervaren uitdaging bekritiseerd. Deze meetinstrumenten waren niet gevalideerd, waren eenzijdig dan wel zeer algemeen of diffuus geoperationaliseerd. Dergelijke conceptualisaties en operationalisaties van ervaren uitdaging kunnen tot onvergelijkbaarheid van en verwarring over onderzoeksresultaten met dit concept leiden. Bij te algemene operationalisaties, blijft bijvoorbeeld altijd de vraag wat de specifieke elementen zijn die een effect verklaren. Dit maakt resultaten lastig te interpreteren. Bij een eenzijdige benadering is de interpretatie eenvoudiger, maar weten we niet hoe andere uitdagingsaspecten zich verhouden tot de gevonden resultaten.

Ons meetinstrument, dat verschillende dimensies behelst, kan in de toekomst wellicht een betrouwbaarder en genuanceerder beeld over de effecten van ervaren uitdaging genereren. Zo werd aangetoond dat competentietoetsing een negatieve relatie heeft met beoordeelde werkprestatie, terwijl er voor de andere schalen geen relatie werd aangetoond. Dit geeft aan dat onze lijst tot divergente resultaten kan leiden. Hiermee wordt aangetoond dat het belangrijk is om de verschillende dimensies van onze lijst mee te nemen in onderzoek naar uitkomsten van ervaren uitdaging. Wellicht kan ons instrument helpen om een beter beeld te krijgen van de relatie tussen uitdaging en werkprestatie waarover nog weinig bekend is, en andere uitkomstvariabelen.

De drie subschalen van ervaren uitdaging zoals in dit onderzoek naar voren zijn gekomen integreren het bestaande onderzoek naar ervaren uitdaging en sluiten aan bij de bestaande literatuur. Ten eerste is de subschaal positieve stimulatie consistent met bestaande operationalisaties waarbij uitdaging wordt geassocieerd met plezierige activerende gemoedstoestanden zoals opwinding, stimulatie en

inzet (e.g., Lazarus & Folkman, 1984; Meyer & Allen, 1988). Ten tweede is de subschaal competentietoetsing consistent met bestaande operationalisaties van taakontwerp waarbij uitdaging wordt geassocieerd met het ontwikkelen van vaardigheden (e.g., Hackman & Oldham, 1976; Jones & James, 1979; Walsh et al., 1980). Ten derde lijkt de subschaal onzekerheid overeen te komen met operationalisaties die in de literatuur over taakkeuze en motivatie worden gebruikt. Zo definieerde McCauley et al. (1999, p. 4) uitdaging onder andere als het moeten oplossen van problemen en het maken van keuzes in risicovolle en onzekere omstandigheden. De bestaande conceptualisaties van uitdaging schoten tekort omdat ze eenzijdig gericht waren op één van de drie dimensies van ervaren uitdaging die in dit onderzoek naar voren zijn gekomen.

Ons onderzoek kende enkele beperkingen die vragen om nader onderzoek. Zo heeft het huidige onderzoek nog niet de predictieve validiteit van ervaren uitdaging kunnen vaststellen. We willen hierbij opmerken dat het onderzoeken van de predictieve validiteit van ervaren uitdaging in het algemeen lastig zal zijn omdat de werkervaringen van mensen aan verandering onderhevig zijn (De Pater et al., 2009; Preenen et al., 2011). Dergelijk voorspellend onderzoek stelt dus bijzondere eisen aan de tijdsperiode waarin predictor en criterium gemeten dienen te worden. Het lijkt raadzaam om die tijdsperiode kort te houden (McDaniel et al., 1988) of meerdere predictor- en criteriummetingen over de tijd te verrichten om op die manier fluctuaties in predictor en criterium te kunnen vaststellen.

Vervolgonderzoek zou ook gebruik moeten maken van meerdere beoordelaars om zodoende het probleem van gedeelde methode variantie te ondervangen (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Deze inzet van meerdere informatiebronnen moet echter weloverwogen gebeuren. Het is bijvoorbeeld maar de vraag of leidinggevenden beter in staat zijn om de commitment van hun medewerker te beoordelen dan de medewerker zelf. Waarschijnlijk niet. Kortom, zelfrapportages leveren soms meer valide gegevens op dan de rapportages van anderen. Wij pleiten aldus voor een combinatie van zelfrapportages en beoordelingen van anderen, daar waar deze een duidelijk toegevoegde waarde hebben. Het lijkt bijvoorbeeld zinvol om ervaren uitdaging van medewerkers te koppelen aan promotiebeoordelingen over medewerkers van leidinggevenden (e.g., DeRue & Wellman, 2009; Dragoni et al., 2009; McCauley et al., 1999).

Hoewel de resultaten van ons onderzoek een eerste ondersteuning bieden voor de betrouwbaarheid en validiteit van het meetinstrument van ervaren uitdaging is aanvullend onderzoek nodig om de relaties tussen ervaren uitdaging en criteriumvariabelen beter te begrijpen. In bestaand onderzoek wordt uitdaging vooral geassocieerd met positieve uitkomsten. De subschalen competentietoetsing en onzekerheid duiden echter op mogelijk negatieve uitkomsten van uitdaging, zoals slechtere werkprestaties en meer stress. Het is van groot belang te onderzoeken welke combinatie van ervaren positieve stimulatie, competentietoetsing en onzekerheid kan zorgen voor positieve effecten voor individu en organisatie.

Tevens zou vervolgonderzoek zich kunnen richten op de externe validiteit van het meetinstrument door bijvoorbeeld te onderzoeken in hoeverre de gevonden resultaten standhouden bij verschillende werksectoren en organisaties.

Toekomstig onderzoek zou eveneens aandacht kunnen besteden aan de antecedenten van ervaren uitdaging, zoals objectieve taakkenmerken en persoonskenmerken. Daarnaast zijn er mogelijke moderatoren in het spel die de relatie tussen taakkenmerken en/of persoonskenmerken en ervaren uitdaging kunnen beïnvloeden (De Pater et al., 2009). Hierbij kan bijvoorbeeld gedacht worden aan de motivationele oriëntaties die mensen in hun werk hebben (Nicholls, 1984). Mensen die doorgaans gericht zijn op het vermijden van falen zullen uitdagende taken waarschijnlijk minder positief waarderen en oppakken (Elliot, 1999) dan mensen die gericht zijn op het nastreven van succes (Hirschfeld, Thomas, & Lankau, 2006). Dergelijke kennis is niet alleen wetenschappelijk interessant, maar ook essentieel voor organisaties die trachten medewerkers aan te trekken, te ontwikkelen en te behouden door ze uitdagend werk aan te bieden. Ons meetinstrument van ervaren uitdaging, het eerste in zijn soort, vormt hiertoe hopelijk een belangrijke aanzet.

Praktijkbox

Wat betekenen de resultaten voor de praktijk?

De uitkomsten suggereren, in lijn met eerder onderzoek, dat ervaren uitdaging een belangrijke factor kan spelen in het verhogen van de werktevredenheid, betrokkenheid en retentie van medewerkers. Organisaties zouden hun medewerkers dus voldoende moeten uitdagen.

De drie subschalen van de ervaren werkuitdaginglijst bieden organisaties empirisch onderbouwde handvatten om de ervaren uitdaging van medewerkers (gericht) te verhogen. Verstrek medewerkers geregeld taken: (a) die hen positief stimuleren, (b) die hun competentie toetsen en waarin men zich moet bewijzen en (c) waarbij men een bepaalde mate van risico, moeilijkheid en onzekerheid ervaart. Het is daarbij vooral belangrijk dat de uitdaging als positief wordt ervaren.

De lijst 'ervaren uitdaging' kan ingezet worden om te peilen in hoeverre medewerkers zich daadwerkelijk uitgedaagd voelen in hun werk, dan wel om interventies ter verhoging van de ervaren uitdaging op een gedegen manier te evalueren.

Samenvatting

Perceived Job Challenge: Development of a measure

A measure of perceived challenge was developed, and its psychometric qualities were investigated in two studies. Perceived challenge consisted of three factors: positive stimulation, competence testing, and uncertainty. These factors were found to be reliable and showed good convergent validity. The non-significant correlations between perceived challenge and the Big Five personality traits agreeableness and neuroticism indicated good discriminant validity. The criterion validity of this measure was good as well. The positive stimulation dimension in particular showed to correlate high with employees' work attitudes. This measure of perceived challenge seems useful for research and practice.

Noten

- 1 Aangezien er meer stellingen betrekking hadden op cognitieve beleving (85 stellingen) dan op affect (28 stellingen).
- 2 De covariantie tussen de residuen van sommige items (respectievelijk items 1, 2, 5, 6, 13 en 14) werden geschat, omdat de modificatie-indices aantoonde dat deze residuen iets gemeenschappelijk hadden, wat te verklaren is door overeenkomst in vorm en/of betekenis van items. De modificatie-indices van de factorladingen waren daarentegen laag.

Literatuur

- Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of Occupational Psychology*, 63, 1-18.
- Behfar, K. J., Peterson, R. S., Mannix, E. A., & Trochim, W. M. K. (2008). The critical role of conflict resolution in teams: A close look at the links between conflict type, conflict management strategies, and team outcomes. *Journal of Applied Psychology*, 93, 170-188.
- Brayfield, A. H., & Rothe, H. F. (1951). An index of job satisfaction. *Journal of Applied Psychology*, 35, 307-311.
- Buchanan, B. (1974). Building organizational commitment: The socialization of managers in work organizations. *Administrative Science Quarterly*, 19, 533-546.
- Cavanaugh, M. A., Boswell, W. R., Roehling, M. V., & Boudreau, J. W. (2000). An empirical examination of self-reported work stress among US managers. *Journal of Applied Psychology*, 85, 65-74.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-98.
- Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, 4, 5-13.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Csikszentmihalyi, M. (1997). *Finding Flow: The psychology of engagement with everyday life*. New York: Basic Books.
- De Gilder, D., Van den Heuvel, H., & Ellemers, N. (1997). Het 3-componenten model van commitment [The three component model of organizational commitment]. *Gedrag & Organisatie*, 10, 95-105.
- De Jonge, J., Dollard, M. F., Dormann, C., Le-Blanc, P. M., & Houtman, I. L. D. (2000). The Demand-Control Model: Specific demands, specific control, and well-defined groups. *International Journal of Stress Management*, 7, 269-287.
- Denison, D. R., Hooijberg, R., & Quinn, R. E. (1995). Paradox and performance: Toward a theory of behavioral complexity in managerial leadership. *Organization Science*, 6, 524-540.
- De Pater, I. E., Van Vianen, A. E. M., Bechtoldt, M. N., & Klehe, U. C. (2009). Employees' challenging job experiences and supervisors' evaluations of promotability. *Personnel Psychology*, 62, 297-325.

- DeRue, D. S., & Wellman, N. (2009). Developing leaders via experience: The role of developmental challenge, learning orientation, and feedback availability. *Journal of Applied Psychology, 94*, 859-875.
- Dixon, M. A., Cunningham, G. B., Sagas, M., Turner, B. A., & Kent, A. (2005). Challenge is key: An investigation of affective organizational commitment in undergraduate interns. *Journal of Education for Business, 80*, 172-180.
- Dragoni, L., Tesluk, P. E., Russell, J. E. A., & Oh, I. S. (2009). Understanding managerial development: Integrating developmental assignments, learning orientation, and access to developmental opportunities in predicting managerial competencies. *The Academy of Management Journal, 5*, 731-743.
- Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist, 34*, 169-189.
- Ettington, D. R. (1998). Successful career plateauing. *Journal of Vocational Behavior, 52*, 72-88.
- Evans, K., & Kersh, N. (2004). Recognition of tacit skills and knowledge. *Journal of Workplace Learning, 16*, 63-74.
- Evers, A. (2001). The revised Dutch rating system for test quality. *International Journal of Testing, 1*, 155-182.
- Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of Applied Psychology, 60*, 159-170.
- Hackman, J., & Oldham, G. (1976). Motivation through the design of work: Test of the theory. *Organizational Behavior and Human Performance, 16*, 250-279.
- Hackman, J. R., & Oldham, G. R. (1980). *Work redesign*. Reading, MA: Addison Wesley Publishing Company.
- Hall, D. T., & Lawler, E. E. (1970). Job characteristics and pressures and the organizational integration of professionals. *Administrative Science Quarterly, 15*, 271-281.
- Hirschfeld, R. R., Thomas, C. H., & Lankau, M. J. (2006). Achievement and avoidance motivational orientations in the domain of mentoring. *Journal of Vocational Behavior, 68*, 524-537.
- Holmes, T. L., & Srivastava, R. (2002). Effects of job perceptions on job behaviors: Implications for sales performance. *Industrial Marketing Management, 31*, 421-428.
- Idsoe, T. (2006). Job aspects in the School Psychology Service: Empirically distinct associations with positive challenge work, perceived control at work, and job attitudes. *European Journal of Work and Organizational Psychology, 15*, 46-72.
- Jones, A. P., & James, L. R. (1979). Psychological climate: Dimensions and relationships of individual and aggregated work environment perceptions. *Organizational Behavior & Human Performance, 23*, 201-250.
- Judge, T. A., Bono, J. E., & Locke, E. A. (2000). Personality and job satisfaction: The mediating role of job characteristics. *Journal of Applied Psychology, 85*, 237-249.
- Jöreskog, K. G., & Sörbom, D. (2006). LISREL 8.80. Chicago: Scientific Software International.
- Kirk-Brown, A., & Wallace, D. (2004). Predicting burnout and job satisfaction in workplace counselors: The influence of role stressors, job challenge, and organizational knowledge. *Journal of Employment Counseling, 41*, 29-37.
- Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.
- Major, D. A., Turner, J. E., & Fletcher, T. D. (2006). Linking proactive personality to motivation to learn and development activity. *Journal of Applied Psychology, 91*, 927-935.
- McCauley, C. D., Ohlott, P. J., & Ruderman, M. N. (1999). *Job Challenge Profile*. New York: Jossey-Bass/Pfeiffer.

- McCaughey, C. D., Ruderman, M. N., Ohlott, P. J., & Morrow, J. E. (1994). Assessing the developmental components of managerial jobs. *Journal of Applied Psychology, 79*, 544-560.
- McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). Job experience correlates of job performance. *Journal of Applied Psychology, 73*, 327-330.
- Meyer, J. P., & Allen, N. J. (1988). Links between work experiences and organizational commitment during the first year of employment: A longitudinal analysis. *Journal of Occupational Psychology, 61*, 195-209.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*, 328-346.
- Podsakoff, N. P., LePine, J. A., & LePine, M. A. (2007). Differential challenge stressor-hindrance stressor relationships with job attitudes, turnover intentions, turnover, and withdrawal behavior: A meta-analysis. *Journal of Applied Psychology, 92*, 438-454.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903.
- Preenen, P. T. Y., De Pater, I. E., Van Vianen, A. E. M. (2008a, August). *Provided job challenge and learning as predictors of turnover intentions, job-search behaviors, and voluntary turnover*. Paper presented at the 2008 conference of the Academy of Management, Anaheim, USA.
- Preenen, P. T. Y., De Pater, I. E., Van Vianen, A. E. M. (2008b, November). *A qualitative analysis of job challenge: A bottom-up conceptualization*. Paper presented at the 2008 conference of the Dutch Association of Researchers in the Field of Work and Organizational Psychology, Heerlen, The Netherlands.
- Preenen, P. T. Y., Van Vianen, A. E. M., De Pater, I. E., & Keijzer, L. (2010). Job challenge, learning, and retention. Manuscript submitted for publication.
- Ragins, B. R., & McFarlin, D. B. (1990). Perceptions of mentor roles in cross-gender mentoring relationships. *Journal of Vocational Behavior, 37*, 321-339.
- Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). *Measures of personality and social psychological attitude*. San Diego: Academic Press.
- Schriesheim, C. A., Powers, K. J., Scandura, T. A., Gardiner, C. C., & Lankau, M. J. (1993). Improving construct measurement in management research: Comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *Journal of Management, 19*, 385-417.
- Schwab, D. P. (1999). *Research Methods for Organizational Studies*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Sheldon, K. M., Ryan, R. M., Rawsthorne, L. J., & Ilardi, B. (1997). Trait self and true self: Cross-role variation in the Big-Five personality traits and its relations with psychological authenticity and subjective well-being. *Journal of Personality and Social Psychology, 73*, 1380-1393.
- Slaughter, J. E., Richard, E. M., & Martin, J. H. (2006). Comparing the efficacy of policy-capturing weights and direct estimates for predicting job choice. *Organizational Research Methods, 9*, 285-314.
- Smith, P. C., Kendall, L. M., & Hulin, C. L. (1969). The measurement of satisfaction in work and retirement: A strategy for the study of attitudes. Chicago: Rand McNally.
- Taylor, M. S. (1981). The motivational effects of task challenge: A laboratory investigation. *Organizational Behavior and Human Performance, 27*, 255-278.
- Tett, R. P., Guterman, H. A., Bleier, A., & Murphy, P. J. (2000). Development and content validation of a 'hyperdimensional' taxonomy of managerial competence. *Human Performance, 13*, 205-251.

- Trochim, W. M. K. (1989). An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning*, 12, 1-16.
- Van Vianen, A. E. M., De Pater, I. E., & Preenen, P. T. Y. (2008). Career Management: Taking control of the quality of work experiences. In J. Athanasou & R. Van Esbroeck (Eds.), *International handbook of career guidance* (pp. 283-301). London: Springer.
- Walsh, J. T., Taber, T. D., & Beehr, T. A. (1980). An integrated model of perceived job characteristics. *Organizational Behavior and Human Performance*, 25, 252-267.
- Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction measures: How good are single-item measures? *Journal of Applied Psychology*, 82, 247-252.
- Wayne, S. J., Shore, L. M., & Liden, R. C. (1997). Perceived organizational support and leader-member exchange: A social exchange perspective. *The Academy of Management Journal*, 40, 82-111.

BOEKBESPREKING

Prestatiemeting van wetenschappelijk onderzoek

Bespreking van M. Pen (2009), Prestatiemeting van wetenschappelijk onderzoek: Een empirische studie naar de doorwerking van universitaire prestatie-meetsystemen (academisch proefschrift Technische Universiteit Delft). Den Haag: Lemma Uitgeverij. 211 pagina's, prijs € 36,50, ISBN 978-90-5931-527-3 (paperback).

De huidige crisis leidt in veel organisaties tot de paradoxale situatie dat er op van alles en nog wat bezuinigd moet worden (inclusief het personeel, vaak de grootste kostenpost), terwijl de hoeveelheid werk niet dienovereenkomstig afneemt – en soms zelfs groeit. Medewerkers moeten dus harder werken en hun tijd efficiënter besteden. Dat is een lastige opgave, zeker als organisaties niet willen of kunnen vervallen in een scientific management-achtige aanpak waarbij de uit te voeren taken tot het uiterste worden geroutiniseerd en het personeel geen enkele zeggenschap meer heeft over hun werk.

Eén sector waarin deze problemen momenteel zeer sterk spelen is die van de Nederlandse universiteiten. Hun budgetten staan zwaar onder druk: met name geld voor onderzoek (de zogenaamde eerste geldstroom) wordt steeds schaarser. Dat vertaalt zich in steeds minder geld voor personeel en eerstegeldstroom-promovendi. Toch wil elke afzonderlijke universiteit haar positie op de Shanghai- dan wel Times-ranking van beste universiteiten behouden dan wel verbeteren en wil Nederland 'kennisland' blijven. Direct daaraan gerelateerd zijn twee lastige problemen. Ten eerste, hoe kan het nog beschikbare onderzoeksgeld het beste worden verdeeld? En ten tweede, hoe kunnen academische onderzoekers gemotiveerd worden om nóg productiever te worden?

Het antwoord op beide vragen is steeds vaker dat universiteiten dan wel faculteiten een vorm van prestatiebeloning zouden moeten invoeren. Wat betreft de vraag wie het weinige geld dat nog voor onderzoek beschikbaar is mag uitgeven, vanuit het oogpunt van prestatiefinanciering is het antwoord daarop dat de beste onderzoekers het meeste recht op deze gelden hebben. Zij 'hebben een beloning voor hun inspanningen verdiend' en 'moeten hun onderzoekslijn kunnen uitbouwen'. Een prettig neveneffect is dat minder succesvol onderzoek op termijn zal verdwijnen, simpelweg omdat de betreffende toebereidende onderzoekers op termijn onvoldoende middelen kunnen genereren om hun onderzoek voort te zetten. Wat betreft de tweede vraag hierboven, prestatieprikkels en de daarmee toenemende concurrentie tussen onderzoekers en onderzoeksgroepen kunnen ervoor zorgen dat de productiviteit toeneemt en dat de kwaliteit van het onderzoek stijgt. Prestatiefinanciering biedt universiteiten dus een op het oog eenvoudige, effectieve

en aantrekkelijke manier om de productiviteit van onderzoekers te bevorderen via de verdeling van onderzoeksgeld.

De vraag die dan rijst is wie die beste onderzoekers zijn. Er zijn allerlei criteria denkbaar: bijvoorbeeld het aantal internationale toppublicaties van een onderzoeker of onderzoeksgroep gedurende een bepaalde periode, het lidmaatschap van een tijdschriftredactie, de verworven subsidies in tweede, derde en soms zelfs vierde (dat wil zeggen, allerlei collectebusfondsen) geldstroom, de mate waarin het werk van een onderzoeker wordt geciteerd door anderen, het aantal afgeleverde promovendi, enzovoort. Gezien de veelheid aan mogelijke criteria en de in potentie grote impact daarvan op onderzoekers is het niet verwonderlijk dat de in- en uitvoering van prestatie-meetsystemen onder professoren veel discussie genereert. Die discussie wordt vaak gevoerd op basis van een mengeling van ideologische ('prestatiebeoordeling bevordert concurrentie en concurrentie bevordert de prestaties'), pragmatische ('ik en mijn groep gaan erop achteruit') en emotionele ('mijn onderzoek is zéker zo goed als het jouwe!') argumenten. Vragen als 'is er een probleem, en wat is dat probleem dan?', 'is prestatiebeoordeling wel de geëigende oplossing voor dat probleem?' en 'levert deze oplossing weer nieuwe moeilijkheden op?' komen bij dit soort discussies gewoonlijk niet aan de orde.

Prestatiemeting van wetenschappelijk onderzoek

Gezien de kwaliteit van deze discussies is het goed dat Mark Pen een proefschrift over de doorwerking van universitaire prestatie-meetsystemen heeft geschreven. Pen was ten tijde van zijn onderzoek werkzaam bij de Koninklijke Nederlandse Akademie van Wetenschappen als stafmedewerker op het gebied van wetenschapsbeleid en onderzoeksmanagement, en als buitenpromovendus verbonden aan de Technische Universiteit Delft. Zijn proefschrift is leerzaam voor (onderzoeks)decanen en -directeuren van universiteiten, maar zeker ook voor in prestatie-meting geïnteresseerde arbeids- en organisatiepsychologen en academische onderzoekers in faculteiten waar gedacht wordt over de invoering van prestatie-meetsystemen.

Achtergrond, ontwikkelingen en vraagstelling. De centrale vraagstellingen in Pens proefschrift zijn (1) welke positieve en negatieve effecten prestatie-meetsystemen in de wetenschappelijke praktijk kunnen hebben, (2) of dergelijke systemen bruikbaar en betekenisvol kunnen zijn voor zowel managers als professionals, en zo ja, (3) onder welke voorwaarden dat het geval zou kunnen zijn.

Pens dissertatie bestaat uit drie delen. In het eerste deel komt de geschiedenis en opmars van prestatie-meting en -financiering aan de orde, alsmede de werking en de veronderstelde voor- en nadelen van prestatie-meetsystemen. Pen legt het begin van de opmars van prestatie-meting aan universiteiten in het midden van de jaren '60 van de vorige eeuw. Publiek en politiek hadden behoefte aan antwoorden op allerlei kennisvragen, en de wetenschap werd geacht daarin te voorzien. Wetenschappelijk onderzoek moest daarom gestuurd en geprogrammeerd worden; daartoe werden adviesraden (zoals de Raad van Advies voor Wetenschapsbeleid in 1966) ingesteld, er kwam een hoofddirectie wetenschapsbeleid op het

ministerie voor Onderwijs en Wetenschappen, en er kwam zelfs een minister-spoot voor wetenschapsbeleid. Die druk van buitenaf vertaalde zich bij universiteiten in een toenemende behoefte aan stuurinformatie in de vorm van output-cijfers, citatieanalyses en onderzoeksevaluaties. Op die manier konden ze hun onderzoeksbeleid en de daarmee samenhangende financiële keuzes zowel intern als extern verantwoorden en laten zien dat hun beleid erop gericht was om de samenhang, maatschappelijke relevantie en kwaliteit van hun onderzoek te bevorderen.

Uiteraard had een en ander ook gevolgen op de werkvloer: denk aan toenemende bureaucratie, verminderde academische vrijheid (geld werd immers toegewezen aan specifieke aandachtsgebieden) en een grotere druk om te presteren (past performance van onderzoekers werd steeds belangrijker bij de verdeling van onderzoeksmiddelen). Het is daarom niet verwonderlijk dat veel onderzoekers huiverig zijn voor prestatiemeting. Onderzoekers opteren vaak voor de al langer bestaande peer review-procedures om de kwaliteit, relevantie en originaliteit van hun onderzoek te evalueren, bijvoorbeeld in de vorm van onderzoeksvisitaties. Deze meer kwalitatieve manier van onderzoeksbeoordeling heeft de laatste jaren concurrentie gekregen van kwantitatieve methoden om onderzoeksprestaties in kaart te brengen: bijvoorbeeld de hoeveelheid extern verworven middelen, de gegenereerde output (het aantal artikelen) en de behaalde effecten van het onderzoek (citaties, octrooien of maatschappelijke toepassingen), maar ook procesmatige indicatoren als het aantal gerealiseerde promoties en het aandeel projectfinanciering. De stap van prestatiebeoordeling naar prestatiefinanciering ligt uiteraard voor de hand. Door een bepaald strategisch gekozen prestatie criterium te benadrukken en de verdeling van middelen daarvan afhankelijk te maken, zullen onderzoekers immers vanzelf het door beleidsmakers gewenste gedrag laten zien. Wil het universitaire management bijvoorbeeld het aantal internationale toppublicaties van medewerkers vergroten, dan kunnen zij de toewijzing van middelen afhankelijk maken van het aantal publicaties in toptijdschriften. Medewerkers zullen vervolgens hun best doen zo veel mogelijk van hun werk in die tijdschriften gepubliceerd te krijgen.

Dit soort simpele sturingsmechanismen kan echter ook strategisch gedrag bij de medewerkers uitlokken (gaming the numbers). Pen geeft daarvan diverse voorbeelden. Het is bijvoorbeeld denkbaar dat als onderzoekers beloond worden op basis van het aantal gerealiseerde publicaties, zij zullen pogen hun data zo dun mogelijk uit te smeren over zo veel mogelijk publicaties (de salamitactiek). Ook zullen zij niet het risico willen lopen dat een onderzoeksproject te weinig output genereert, waardoor vernieuwend (en dus risicovol) onderzoek niet meer uitgevoerd wordt (afname van creativiteit). Ten slotte zullen onderzoekers hun prioriteiten gaan verleggen: als alleen publicaties in internationale toptijdschriften tellen bij het bepalen van onderzoeksprestaties, wat is dan nog het nut van het schrijven van Nederlandstalige artikelen, hoofdstukken en boekbesprekingen? En waarom zou je je tijd verdoen aan redactiewerk of het bijhouden van je vak in bredere zin (dus ook eens een boek lezen dat niet direct relevant is voor je onderzoek) als je daarvoor door je werkgever wordt gekort op je onderzoekstijd vanwege verminderde productiviteit (deprofessionalisering)? De mate waarin dit

soort (in Pens woorden) pervers gedrag een probleem vormt zal uiteraard afhangen van de criteria op basis waarvan prestatie wordt gekwantificeerd. Pen besluit dit deel van zijn dissertatie dan ook met een opsomming van de aspecten waaraan een zinvol prestatie-meetsysteem moet voldoen. Belangrijk is bijvoorbeeld dat het systeem legitiem is, dat wil zeggen dat het door betrokkenen wordt geaccepteerd als een systeem dat haalbare doelen stelt, prestaties belooft die passen bij de aard van het vakgebied (dus niet focust op bijvoorbeeld het aantal verkregen octrooien als een wetenschapsgebied zich daarvoor niet leent) en inzicht verschaft in de doeltreffendheid en doelmatigheid van het functioneren van de organisatie. Ook moet een meetsysteem rekening houden met de grote variëteit aan wetenschappelijke producten; er moet dus een dienovereenkomstige mix aan indicatoren zijn, wat impliceert dat meetsystemen in de praktijk vaak (zeer) ingewikkeld zullen zijn. Ten slotte moet het systeem dynamisch zijn. Het verliest zijn functionaliteit op het moment dat iedereen (of niemand) aan de maatstaven voldoet; het systeem kan dus niet in beton worden gegoten, maar moet bijgesteld kunnen worden wanneer dat nodig is.

De praktijk. In het tweede deel van zijn dissertatie bespreekt Pen vier casussen die de gang van zaken beschrijven bij vier faculteiten (technische natuurwetenschappen, geneeskunde, economische wetenschappen en sociale wetenschappen) van drie Nederlandse universiteiten. In elk van deze casussen is sprake van een al langer functionerend prestatie-meetsysteem, zodat de effecten daarvan op budgetontwikkeling en onderzoeksprestaties alsmede eventuele problemen zichtbaar zouden moeten zijn. Er werden zowel documenten (jaarverslagen, formatiecijfers, publicatielijsten) bestudeerd als interviews gehouden met ruim vijftig decanen, afdelingshoofden (hoogleraren) en beleidsmedewerkers. Op die manier werd gepoogd om ook de ervaringen in het veld met de systemen te documenteren.

Hoewel de systemen inhoudelijk sterk variëren over de casussen, blijkt het effect ervan op de verdeling van het budget over onderzoekers en hun groepen niet groot. Eén belangrijke reden daarvoor is dat vaak slechts een beperkt deel van het eerstegeldstroom-budget dynamisch over groepen wordt verdeeld. Gemeenschappelijke voorzieningen als bibliotheek, huisvesting, laboratoria, centrale staf en dergelijke komen immers vaak ten laste van dat budget, maar kunnen niet of nauwelijks prestatieafhankelijk worden toegekend. Ook beschikt de decaan gewoonlijk over een eigen budget voor strategisch onderwijs- en onderzoeksbeleid, dat evenmin wordt verdeeld op basis van prestaties van de groepen. Een tweede belangrijke reden is volgens Pen dat onderzoeksgroepen en onderzoekers ervoor kunnen kiezen om de prestatiecriteria waaraan voldaan moet worden om aanspraak te kunnen maken op fondsen vanuit de eerste geldstroom te negeren, en zich in plaats daarvan te richten op financiering vanuit de tweede en vooral derde geldstroom. Het budget voor onderzoek kan zo gelijk blijven of zelfs toenemen, zonder dat men hoog scoort op door de universitaire bestuurders gedefinieerde prestatiecriteria. Een derde belangrijke reden voor het geringe effect van prestatie-meetsystemen op de verdeling van budgetten over onderzoekers en groepen is dat universitaire bestuurders vaak bewust het effect van prestatie-meetsystemen op de verdeling van onderzoeksgelden proberen te dempen. Decanen kunnen bijvoorbeeld de eigen middelen gebruiken om eventuele zeer negatieve effecten van

prestatiebeloning voor een bepaalde groep te compenseren. Het is bijvoorbeeld denkbaar dat een groep nauwelijks internationaal publiceert (één van de belangrijkste criteria in vrijwel alle meetsystemen) maar wel veel opdrachtonderzoek binnenhaalt, daarmee een belangrijke maatschappelijke functie heeft en dus verdient om te blijven bestaan. Een laatste belangrijke reden is dat prestatiefinanciering er weliswaar toe kan leiden dat productiviteit en kwaliteit van onderzoek toenemen, maar als die stijging voor alle onder het prestatiemeetsysteem vallende groepen ongeveer even groot is zal er uiteindelijk geen verschuiving in de budgetten te zien zijn.

In de praktijk blijken zowel het management als afdelingshoofden dit soort systemen eerder te zien als een middel om discussies over de verdeling van onderzoeksgelden te voorkómen (het systeem geeft daar immers objectieve en idealiter breed gedragen regels voor), dan om prestaties te bevorderen. Dat kan ertoe leiden dat dergelijke systemen behoorlijk complex worden. Universitair onderzoek kan immers leiden tot een veelheid aan wetenschappelijke output (wetenschappelijke publicaties, editorships, octrooien, boekhoofdstukken, proceedings, onderzoeksvoorstellen, promoties, rapporten aan externe organisaties, publieksvoorlichting, optredens in de media, managementtaken, enzovoort), en al dit soort output – voor zover waardevol bevonden – kan in potentie verdisconteerd worden in het prestatiemeetsysteem. Omdat het systeem vaak de uitkomst is van een onderhandelingsproces wordt die potentiële complexiteit ook vaak bewaarheid; iedere partij wil immers zijn eigen favoriete indicator in het meetsysteem terugzien. In de praktijk blijkt het niet gemakkelijk een meetsysteem te ontwikkelen dat recht doet aan alle relevante wetenschappelijke output en de tijdsinvesteringen die daarmee gepaard gaan (een editorship kost bijvoorbeeld vaak veel tijd en draagt bij aan het wetenschappelijke prestige van een groep, maar wordt zelden in prestatiemeetsystemen opgenomen).

De besproken casussen laten zien dat de effecten van prestatiemeetsystemen op het gedrag van medewerkers vaak nogal beperkt zijn. Dat is niet zo verwonderlijk: onderzoekers hebben immers vaak sowieso al een sterke intrinsieke motivatie om goed onderzoek te doen en zo veel mogelijk te scoren in goede tijdschriften. Wel blijken onderzoekers zich sterker te gaan richten op het type output dat door het systeem beloond wordt, zij publiceren bijvoorbeeld vaker in Engelstalige tijdschriften als dat door het meetsysteem wordt beloond. Als er al sprake is van een productiviteitsstijging, is dat vaak te danken aan het vervangingseffect: oudere, minder productieve onderzoekers worden na pensionering vervangen door jongere onderzoekers die beter scoren op de criteria van het meetsysteem. Uit de casussen blijkt verder dat onderzoeksbudgetten maar marginaal veranderen als functie van de invoering van prestatiefinanciering, vooral om de al eerder genoemde redenen. Onmiddellijk na invoering blijkt het effect iets groter te zijn (de zwakke broeders in een groep, dan wel de zwakke groepen krijgen al snel te maken met een afname van hun budget), maar na verloop van tijd variëren onderzoeksbudgetten nauwelijks meer.

Al met al zijn vooral managers gelukkig met de onderzochte prestatiemeetsystemen. Anderzijds zijn er maar weinig onderzoekers écht negatief over deze systemen, wellicht juist omdat de effecten ervan op de verdeling van onderzoeksgelden

in de praktijk niet zo heel groot zijn. De bijdrage van dit soort systemen lijkt vooral te bestaan uit het feit dat ze duidelijk maken welk prestatieniveau er van onderzoekers wordt verwacht, waardoor het kan bijdragen aan een prestatiege-richte cultuur op de betreffende faculteiten. Bovendien kan met dit soort syste- men het onderzoeksgeld volgens relatief objectieve criteria verdeeld worden. Nadelen zijn dat de rechtvaardigheid en compleetheid van die criteria steeds onderwerp van discussie zijn, dat het systeem kan leiden tot veel tijdverlies van- wege de toegenomen administratieve belasting, en dat het kan leiden tot interne strijd tussen afdelingen. Prestatiefinanciering is immers in principe een zero-sum game: de totale hoeveelheid eerstegeldstroom-geld neemt eerder af dan toe, wat betekent dat de winst van de één automatisch tot verlies bij de ander leidt. Dat is vooral onwenselijk en frustrerend als de verschillen in productiviteit en kwaliteit tussen groepen maar klein zijn. Het kan ook leiden tot inefficiëncy en ondoelma- tigheid; de strijd tussen onderzoekers en groepen onderling kost tijd en energie die wellicht vruchtbaarder zou zijn besteed aan het schrijven van topartikelen, dan wel aan het verwerven van fondsen in de tweede en derde geldstroom.

Ontwerpprincipes voor prestatie-meetsystemen. In het derde deel (dat feitelijk uit maar één hoofdstuk bestaat) trekt Pen kort conclusies en worden er aanbeve- lingen ten behoeve van het ontwerp van prestatie-meetsystemen geformuleerd. Dat prestatie-meetsystemen in potentie meer kwaad dan goed kunnen doen is dui- delijk. Nadelige effecten hebben bijvoorbeeld betrekking op onderzoekers die strategisch gaan handelen en bepaalde nuttige en zinvolle taken gaan verwaarlo- zen ten faveure van bezigheden die beter scoren op de gekozen prestatiecriteria, toegenomen bureaucratie die niet leidt tot prestatieverbetering of een vergroting van de te verdelen middelen, en overmatig complexe meetsystemen die door de betrokken onderzoekers nog steeds als onrechtvaardig worden gezien. Om de positieve effecten te laten prevaleren, moeten dergelijke systemen aan een aantal factoren voldoen: de prestatie-indicatoren moeten legitiem zijn en door de betrokkenen als zodanig worden geaccepteerd, alle relevante activiteiten en resul- taten moeten erin terug te vinden zijn, en de systemen moeten kunnen worden aangepast aan veranderingen in de omgeving.

Pen stelt dat er twee belangrijke groepen succesfactoren zijn: dempingsfactoren en compenserende factoren. Dempingsfactoren zorgen ervoor dat de effecten van het prestatie-meetsysteem minder impact hebben dan het geval zou zijn als de dempingsfactoren niet aanwezig zouden zijn. Denk bijvoorbeeld aan vertraagde doorwerking van effecten (prestaties in jaar x hebben pas in jaar $x+1$ of nog later effect op de toewijzing van middelen; bij de toewijzing van middelen kunnen de prestaties over een langere periode meespelen, waardoor al te grote schokeffecten uitblijven), of aan uitzonderingsposities voor onderzoekers dan wel bepaalde typen output (bijvoorbeeld omdat die in het meetsysteem ondergewaardeerd wor- den). Compenserende factoren zorgen ervoor dat de negatieve effecten van het prestatie-meetsysteem opgevangen worden door andere maatregelen. Het kan dan bijvoorbeeld gaan om alternatieve manieren om de middelen uit de eerste geld- stroom te verdelen die parallel aan prestatiefinanciering worden gehanteerd. Bovendien wordt veel geld via de tweede en derde geldstroom verworven, wat de mogelijke negatieve effecten van het prestatie-meetsysteem verder zal verzachten.

De aanwezigheid van deze twee typen succesfactoren zorgt ervoor dat onderzoekers de prikkels die uitgaan van het prestatiebeloningssysteem in (veel) mindere mate zullen voelen. Pen trekt daarmee impliciet de nogal paradoxale conclusie dat het succes van prestatie-meetsystemen vooral afhangt van de mate waarin de impact van prestatie-meting op de verdeling van middelen uit de eerste geldstroom beperkt wordt.

Evaluatie: gevolgen voor onderzoekers

Pens proefschrift geeft een interessant beeld van de manier waarop universiteiten, managers en onderzoekers worstelen met prestatie-meting en -financiering. Met name de theoretische analyse en de beschrijving van de casussen zijn fascinerend en herkenbaar. De analyses en conclusies zijn soms minder overtuigend, wat deels ook samenhangt met het gekozen onderzoeksontwerp. Onduidelijk blijft bijvoorbeeld in hoeverre de geconstateerde productiviteitsstijging te danken is aan de besproken prestatiebeloningssystemen. Dát er in alle casussen sprake was van een productiviteitsstijging valt niet te ontkennen, maar omdat de Nederlandse wetenschap als geheel in de onderzochte periode eveneens beter is gaan scoren (dus ook de onderzoekers die niet te maken hadden met prestatie-meetsystemen), kan betwijfeld worden of prestatie-financiering hierbij de doorslaggevende factor was. Evenmin is duidelijk of de kwantitatieve manier van prestatie-meting en -financiering anders of beter werkt dan het traditionele peer-to-peer beoordelings-systeem zoals dat in onderzoeksvisitaties wordt toegepast.

Weer een ander punt betreft de reden waarom de deelnemende faculteiten overgingen tot invoering van een systeem van prestatie-financiering. Welk onderliggend probleem werd hiermee aangepakt? Het is bijvoorbeeld denkbaar dat het systeem werd ingevoerd om een improductieve en vastgeroeste faculteit wakker te schudden en zwakke, improductieve onderzoekers te lozen. Prestatie-financiering kan in dat soort omstandigheden inderdaad een effectieve oplossing zijn. Het is echter een open vraag of dit soort systemen in andere contexten – bijvoorbeeld als er geen duidelijk zwakke broeders zijn – net zo effectief zal zijn, dan wel alleen maar negatieve effecten teweeg zal brengen (veel administratief gedoe en interne strijd, terwijl de uiteindelijke verschuivingen in het onderzoeksbudget maar marginaal zijn).

Ten slotte is het om twee redenen denkbaar dat het proefschrift van Pen een alweer wat verouderd beeld geeft van het type criteria dat in prestatie-meetsystemen is opgenomen. Ten eerste, de nadruk in die systemen ligt vaak op output, maar nu de eerste geldstroom opdroogt en de competitie om gelden uit tweede en derde geldstroom steeds intensiever wordt is het aannemelijk dat de werfkracht van onderzoekers een steeds grotere rol zal krijgen bij het beoordelen van hun prestaties. Ten tweede, prestatiebeoordelingssystemen zijn blijkens Pens proefschrift vaak in ieder geval ten dele bedoeld om de rust onder de staf te bevorderen door onderzoekstijd volgens vaste en duidelijke regels te verdelen. Er zijn echter omstandigheden denkbaar waaronder dat soort regels slecht uitpakt. Om een VIDI-laureaat van buiten aan te trekken kan het bijvoorbeeld noodzakelijk zijn

om die persoon meer onderzoekstijd te geven dan volgens het vigerende prestatiebeoordelingssysteem gerechtvaardigd zou zijn. Omdat dit soort situaties de laatste jaren vaker lijkt voor te komen, zal het in de toekomst vermoedelijk belangrijker worden om arbeidsvoorwaarden af te kunnen stemmen op de persoon in kwestie. De huidige, vaak tamelijk rigide prestatiebeoordelingssystemen zijn daar niet voor bedoeld en werken die noodzakelijke flexibiliteit vaak tegen. Kortom, hoewel Pens dissertatie veel stof tot nadenken biedt, is het laatste woord over de effecten en de bruikbaarheid van prestatiefinanciering van onderzoek bepaald nog niet gezegd.

Toon Taris

Universiteit Utrecht, afdeling Arbeids- en organisatiepsychologie

REVIEWERS

Jaargang 2010

De redactie bedankt de onderstaande reviewers hartelijk voor het beoordelen van manuscripten in 2010.

Dr. Wieby Altink
Prof. dr. Arnold Bakker
Dr. Matthijs Bal
Dr. Debby Beckers
Prof. dr. Marise Born
Dr. Wim van Breukelen
Dr. Lieke ten Brummelhuis
Dr. Karen van Dam
Dr. Daantje Derks
Dr. Eva Deros
Dr. Marloes van Engen
Prof. dr. Sabine Geurts
Drs. Marije de Goede
Dr. Marjan Gorgievski
Prof. dr. Beate van der Heijden
Dr. Astrid Homan
Dr. Laurens ten Horn
Dr. Inge Houkes
Suzanne Jak M.Sc.
Dr. Onne Janssen
Dr. Rendel de Jong
Prof. dr. Jan de Jonge
Prof. dr. Michiel Kompier
Dr. Dorien Kooij
Dr. Jonas Lang
Dr. Annet de Lange
Dr. Dimitri van der Linden
Dr. Karen Nieuwenhuijsen
Gera Noordzij M.Sc.
Prof. dr. Bernard Nijstad
Prof. dr. Karen van Oudenhoven-van der Zee
Dr. Maria Peeters
Mirjam Radstaak M.Sc.
Prof. dr. René Schalk
Dr. Judith Sluiter
Dr. Omar Solinger

Ellen Spoor M.Sc.
Dr. Herman Steensma
Prof. dr. Toon Taris
Prof. dr. Mandy van der Velde
Dr. Bart Verkuil
Dr. Peter Vlerick
Prof. dr. René van der Vlist
Drs. Anita de Vries
Dr. Reinout de Vries
Prof. dr. Tinka van Vuuren
Alfred Wagenaar M.Sc.
Dr. Jan Fekke Ybema
Prof. dr. Fred Zijlstra

AUTEURSREGISTER

Jaargang 2010

Baas, Matthijs, Carsten K.W. De Dreu en Bernard A. Nijstad, Stemming, motivationele oriëntatie, fit en creativiteit: De rol van mentale activering 73

Bakhuys Roozeboom, Maartje C. , Zie Hooff, Madelon L.M. van, Maartje C. Bakhuys Roozeboom , Ernest M.M. de Vroome en Peter G.W. Smulders

Bal, P. Matthijs , Annet H. de Lange, Paul G.W. Jansen en Mandy E.G. Van der Velde, Leeftijd, het psychologisch contract en werkattitudes: een meta-analyse 44

Bloom, Jessica de, Sabine Geurts en Michiel Kompier, Vacation from work as prototypical recovery opportunity 333

Born, Marise Ph., Het beoordelen van mensen voor werkgerelateerde doelen: Etniciteit, moraliteit, en subjectiviteit 232

Dreu, Carsten K.W. De, Een kwantitatieve analyse van Nederlandse bijdragen aan internationaal arbeids- en organisatiepsychologisch onderzoek 28

Dreu, Carsten K.W. De, Zie Matthijs Baas, Carsten K.W. De Dreu en Bernard A. Nijstad

Evers, Arne, Zie Jak, Suzanne en Arne Evers

Geurts, Sabine, Zie Bloom, Jessica de, Sabine Geurts en Michiel Kompier

Görtz, Saskia, Zie Nauta, Aukje, Annet H. de Lange en Saskia Görtz

Hamers, Jan P.H., Zie Spoor, Ellen, Jan de Jonge en Jan P.H. Hamers

Hooff, Madelon L.M. van, Maartje C. Bakhuys Roozeboom, Ernest M.M. de Vroome en Peter G.W. Smulders, De invloed van afwijkende werktijden op de werk-thuis situatie 175

Houkes, Dr. Inge, Zie Verdonk, Dr. Petra en Dr. Inge Houkes

Houtman, Irene L.D., Zie Vroome, Ernest M.M. de, Peter G.W. Smulders en Irene L.D. Houtman

Hülshager, Ute R., Zie Walkowiak, Alicia L.T., Ute R. Hülshager en Fred R.H. Zijlstra

Jacobs, Mardou, Zie Vrugt, Anneke, Mardou Jacobs En Roel Meertens

Jak, Suzanne en Arne Evers, Onderzoeksnotitie. Een vernieuwd meetinstrument voor organizational commitment 158

Jansen, Paul G.W., Zie Bal, P. Matthijs, Annet H. de Lange, Paul G.W. Jansen en Mandy E.G. Van der Velde

Jonge, Jan de, Fred R.H. Zijlstra en Judith K. Sluiter, Nu even niet...! Over herstel en herstelstrategieën bij arbeid 259

Jonge, Jan de, Zie Spoor, Ellen, Jan de Jonge en Jan P.H. Hamers

Kompier, Michiel, Zie Bloom, Jessica de, Sabine Geurts en Michiel Kompier

Lange, Annet H. de, Zie Bal, P. Matthijs, Annet H. de Lange, Paul G.W. Jansen en Mandy E.G. Van der Velde

Lange, Annet H. de, Zie Nauta, Aukje, Annet H. de Lange en Saskia Görtz

Lievens, Filip, Zie Schollaert, Eveline en Filip Lievens

Meertens, Roel, Zie Vrugt, Anneke, Mardou Jacobs en Roel Meertens

Nauta, Aukje, Annet H. de Lange en Saskia Görtz, Lang zullen ze leven, werken en leren. Een schema voor het begrijpen en beïnvloeden van inzetbaarheid gedurende de levensloop 136

Nijstad, Bernard A., Zie Baas, Matthijs, Carsten K.W. De Dreu en Bernard A. Nijstad

Peters, Pascale en Marijn Wildenbeest, Telewerken als hulpbron? 'Flow' en uitputting onder twee telewerkcategorieën vergeleken 97

Ruyssveldt, Joris Van en John Taverniers, Al werkend leren? De actief leren-hypothese van Karasek revisited 1

Ruyssveldt, Joris Van en John Taverniers, Werkplekieren en emotionele uitputting in het Job Demands-Resources-model 118

Schollaert, Eveline en Filip Lievens, Een andere invalshoek op de rollenspelertaining in Assessment Centers 213

Sluiter, Judith K., Zie Jonge, Jan de, Fred R.H. Zijlstra en Judith K. Sluiter

Sluiter, Judith K., Zie Veldhoven, Marc J.P.M. van en Judith K. Sluiter

Smulders, Peter G.W., Zie Hooff, Madelon L.M. van, Maartje C. Bakhuys Roozeboom, Ernest M.M. de Vroome en Peter G.W. Smulders

Smulders, Peter G.W., Zie Vroome, Ernest M.M. de, Peter G.W. Smulders en Irene L.D. Houtman

Spoor, Ellen, Jan de Jonge en Jan P.H. Hamers, Nu even niet...! Of toch wel...? Een dagboekstudie naar detachment en creativiteit 296

Taverniers, John, Zie Ruyssveldt, Joris Van en John Taverniers

Velde, Mandy E.G. Van der, Zie Bal, P. Matthijs, Annet H. de Lange, Paul G.W. Jansen en Mandy E.G. Van der Velde

Veldhoven, Marc J.P.M. van en Judith K. Sluiter, Herstelmogelijkheden en gezondheid in het werk 275

Vroome, Ernest M.M. de, Peter G.W. Smulders en Irene L.D. Houtman, Longitudinale studie naar oorzaken en effecten van presenteïsme 194

Vroome, Ernest M.M. de, Zie Hooff, Madelon L.M. van, Maartje C. Bakhuys Roozeboom, Ernest M.M. de Vroome en Peter G.W. Smulders

Vrugt, Anneke, Mardou Jacobs en Roel Meertens, Reductie van negatieve stereotypen over Marokkanen door een speelfilm 19

Walkowiak, Alicia L.T., Ute R. Hülshager en Fred R.H. Zijlstra De relatie tussen herstel, werkdruk en slaap: een dagboekstudie 316

Wildenbeest, Marijn, Zie Peters, Pascale en Marijn Wildenbeest

Zijlstra, Fred R.H., Zie Jonge, Jan de, Fred R.H. Zijlstra en Judith K. Sluiter

Zijlstra, Fred R.H., Zie Walkowiak, Alicia L.T., Ute R. Hülshager en Fred R.H. Zijlstra

SUMMARIES

Recruitment & Selection: Back from never been gone

Eva Derous, Mandy E.G. van der Velde & Marise Ph. Born

E. Derous, M. E. G. van der Velde & M. Ph. Born (2011). *Gedrag & Organisatie*, volume 24, January 2011, nr. 1, pp. xx-xx. The Netherlands and Flanders are well known for their long and profound tradition in research and development on recruitment and selection practices. During the past decade, however, considerably less attention has been paid towards recruitment and selection topics, both within the Dutch/Flemish academic literature as well as at Dutch/Flemish universities (e.g., in terms of the number of specialized courses). This remarkable observation runs counter to the call for scientifically trained recruiting/selection experts as instigated by several new developments within the professional field of recruitment and assessment. This article presents six developments/trends in recruitment and selection. We further illustrate the scientific-practitioner gap as well as several perspectives on recruitment and selection to address some of the recent developments.

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

Filip Lievens & Britt De Soete

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 conditional 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.

Wilmar Schaufeli, Corine van Wijnhe, Maria Peeters & Toon Taris

Workaholism or work addiction is defined as the uncontrollable need to work excessively hard. Based on this definition, the DUWAS (Dutch Workaholism Scale) is introduced, a brief two-dimensional self-report instrument that includes two correlated subscales of five items each: Working Excessively (WE) and Working Compulsively (WC). Using a national Dutch database ($n = 11,060$) cut-off scores were computed that allow a classification of levels of workaholism. Analyses ($n = 1,166$) reveal that particularly the combination of high scores on both WE and CW is unfavorable. It is concluded that the DUWAS is a reliable and valid inventory that can, for instance, be used as a screening tool to identify workaholic employees.

Perceived Job Challenge: Development of a measure

Paul T. Y. Preenen,

Annelies E. M. van Vianen, Irene E. de Pater & René Geerling

A measure of perceived challenge was developed, and its psychometric qualities were investigated in two studies. Perceived challenge consisted of three factors: positive stimulation, competence testing, and uncertainty. These factors were found to be reliable and showed good convergent validity. The non-significant correlations between perceived challenge and the Big Five personality traits agreeableness and

Summaries

neuroticism indicated good discriminant validity. The criterion validity of this measure was good as well. The positive stimulation dimension in particular showed to correlate high with employees' work attitudes. This measure of perceived challenge seems useful for research and practice.

Hoe zet je een auditsysteem op?

Kwaliteitsmanagement en interne auditing

Bernadette van Pampus



Kwaliteitsmanagement is het doorlopend verbeteren van interne processen. Regelmatig moet worden getoetst of het systeem nog doeltreffend is of dat er verbeteringen nodig zijn. De interne audit is een onderzoek om zwakke plekken in het kwaliteitsmanagementsysteem op te sporen.

In dit boek komt aan de orde hoe het proces van de interne audit verloopt, wat verwacht kan worden van de auditor en wat er gebeurt met de resultaten. Het boek bevat verder praktische tips en suggesties om een auditsysteem op te zetten en te onderhouden. Talloze praktijkvoorbeelden laten zien wat de winsten, maar ook wat de valkuilen van een interne audit kunnen zijn.

Prijs € 23,00

inclusief btw,
exclusief verzend- en
administratiekosten,
ook verkrijgbaar
via de boekhandel
ISBN 978-90-473-0107-3

BESTEL NU VIA WWW.BOOMONDERWIJS.NL

Boom onderwijs

Postbus 85576 | 2508 CG Den Haag | telefoon 070 330 70 33 | fax 070 330 70 30
e-mail info@boomonderwijs.nl | website www.boomonderwijs.nl

Integrale organisatievernieuwing

Slimmer werken en sociale innovatie

Peter Oeij, Luc Dorenbosch, John Klein Hesselink en Fietje Vaas



Prijs € 25,00

inclusief btw,
exclusief verzend- en
administratiekosten,
ook verkrijgbaar
via de boekhandel
ISBN 978-90-5931-570-9

Dit boek plaatst de begrippen slimmer werken en sociale innovatie in een theoretisch kader dat is gestoeld op managementwetenschap, bedrijfskunde en HRM. Van daaruit belichten de auteurs resultaten uit empirisch onderzoek onder Nederlandse bedrijven en instellingen: hoe staan zij ervoor met slimmer werken en sociale innovatie? Ook worden praktijkvoorbeelden van slimmer werken en sociale innovatie beschreven op basis van succesvolle 'cases'. Daarna behandelen de auteurs het veranderkundig ontwikkelen en invoeren van interventies met slimmer werken en sociale innovatie. Hiertoe wordt een stappenplan aangeboden en een overzicht van mogelijke interventies.

Het boek is bedoeld voor managers, beleidsmakers, onderzoekers, adviseurs en studenten die meer willen weten over slimmer werken en sociale innovatie.

BESTEL NU VIA WWW.BOOMLEMMAUITGEVERS.NL

BOOM | LEMMA
UITGEVERS

Postbus 85576 | 2508 CG Den Haag | telefoon 070 330 70 33 | fax 070 330 70 30
e-mail verkoop@budh.nl | website www.boomlemmauitgevers.nl