

**Virtual Reality Assessment:
De Ontwikkeling van een Innovatieve Situationele Beoordelingstest
voor Leidinggeven**



Rijsdijk, Joey
juni 2019

Supervisor: Marise Born
Tweede beoordelaar: Annemarie Hiemstra
Instellingsbegeleider: Barend Koch
Erasmus School of Social and Behavioural Sciences
Erasmus Universiteit Rotterdam

Samenvatting

In dit onderzoek staat de ontwikkeling van een *virtual reality* (VR) situationele beoordelingstest of zogeheten *situational judgment test* (SJT) centraal. De VR-SJT beoogt het inzicht in effectief leiderschap te meten en rapporteert naast een totaalscore iemands scores op taak- en persoonsgericht leidinggeven. Het doel van dit onderzoek was het bepalen van de psychometrische eigenschappen van het instrument. Daartoe werd de betrouwbaarheid en de begripsvaliditeit van het instrument onderzocht. Daarnaast werd de soortgenootvaliditeit met een video-SJT met een vergelijkbare meetpretentie bestudeerd. Verder werd onderzocht hoe de VR-SJT verband houdt met cognitieve capaciteiten en persoonlijkheid en werd de ervaring van deelnemers bestudeerd. Onderzoek onder 101 deelnemers liet zien dat de VR-SJT een betrouwbaar instrument is om inzicht in effectief leiderschap te meten. De convergente en discriminante validiteit werd aangetoond voor de domeinen taakgericht en persoonsgericht leidinggeven. De VR-SJT vertoonde geen relatie met cognitieve capaciteiten en een onverwachte negatieve relatie met extraversie. Tot slot werd gevonden dat deelnemers de VR-SJT positiever beoordelen dan een video-SJT. Zij beoordeelden de VR-SJT onder andere als leuk, realistisch, modern en relevant voor leidinggevende vaardigheden. Ook de testmotivatie was erg hoog. Beperkingen van dit onderzoek en suggesties voor toekomstig onderzoek naar de VR-SJT worden besproken. Ten slotte wordt gereflecteerd op de waarde van dit onderzoek en de implicaties voor de praktijk.

Publiekssamenvatting

Tegenwoordig wordt er door veel werkgevers gebruik gemaakt van psychologische tests om (toekomstige) werknemers te toetsen, selecteren, trainen en beoordelen. Hoewel bekende tests zoals de cognitieve capaciteitentest en de persoonlijkheidsvragenlijst uitvoerig zijn onderzocht, spreken zij weinig tot de verbeelding van zowel organisaties als van deelnemers. Daarnaast lijken dit soort instrumenten op het eerste oog weinig raakvlakken te hebben met daadwerkelijke werkzaamheden. Door recentelijke technologische ontwikkelingen, zoals *virtual reality* (VR), kunnen tests aantrekkelijker en meer realistisch worden gemaakt. In dit onderzoek is dan ook gekeken naar de inhoudelijke eigenschappen van een situationele beoordelingstest (SJT) in VR format. De VR-SJT is ontwikkeld om inzicht in effectief leiderschap te meten en neemt de kandidaat mee naar een virtuele vergaderomgeving waar zich verschillende gebeurtenissen

afspelen. De kandidaat wordt gevraagd om het gedrag van een leidinggevende te beoordelen op effectiviteit. De antwoorden worden vergeleken met een groep ervaren leidinggevers en andere experts. Na afloop van de test wordt gerapporteerd op het totale inzicht in effectief leiderschap en op de domeinen taak- en persoonsgericht leidinggeven. In dit onderzoek wordt gevonden dat de VR-SJT een betrouwbare manier is om inzicht in effectief leiderschap te meten. De VR-SJT wordt positief ontvangen door kandidaten, die de test bovendien positiever beoordelen dan een video-SJT die hetzelfde meet. Kandidaten beoordeelden de VR-SJT als leuk, realistisch, modern en relevant voor leidinggevende vaardigheden. Ook de motivatie om de test te maken was erg hoog. De VR-SJT vormt een innovatieve aanwinst voor assessmentbureaus en andere HR-professionals en is tevens een manier om zich te onderscheiden van de concurrentie en om de selectieprocedure aantrekkelijker te maken voor kandidaten.

Virtual Reality Assessment: De Ontwikkeling van een Innovatieve Situationele Beoordelingstest voor Leidinggeven

Het selecteren van personeel behoort van oudsher tot een van de belangrijkste en meest kenmerkende domeinen binnen de arbeids- en organisatiepsychologie. Menig arbeidspsycholoog heeft dan ook uitvoerig kennis kunnen nemen van het uitgebreide onderzoek dat is verricht naar zaken zoals de betrouwbaarheid en validiteit van traditionele selectiemethoden zoals het gestructureerde interview en de cognitieve capaciteitentest (Schmidt & Hunter, 1998). Deze selectiemethoden behoren tot een van de meest onderzochte methoden in het veld (Schmidt & Hunter, 1998).

Sinds de drastische toename van het internet begin jaren 90 en onze constant ontwikkelende behoefte om zowel ons privé als ons professionele leven een digitale draai te geven, heeft er ook in het domein van personeelselectie veel verandering plaatsgevonden en zijn veel selectiemethoden gedigitaliseerd en/of op andere wijze vernieuwd. Er wordt inmiddels gebruik gemaakt van computer- en internetgesteunde tests, er worden er interviews op afstand gehouden via bijvoorbeeld Skype, en video's, webcams en andere multimediale tools worden ingezet om een kandidaat te toetsen en te selecteren (Anderson, 2003; Lievens & De Soete, 2011). Zulk soort vernieuwende en innovatieve selectiemethoden zijn tevens een manier voor organisaties om zich te onderscheiden; *employer branding* is belangrijk geworden om toekomstige werknemers te werven (Backhaus & Tikoo, 2004; Lievens, 2007). In een schaarse arbeidsmarkt is het belangrijk voor organisaties om zich zo aantrekkelijk mogelijk te verkopen aan mogelijk talent. Heineken heeft bijvoorbeeld recentelijk een spraakmakende interactieve online sollicitatieprocedure opgezet, waar echte medewerkers van Heineken de sollicitant kennis laten maken met de bedrijfscultuur door middel van een luchtige bedrijfsfilm gekoppeld aan een reeks interviewvragen. Met deze campagne wil Heineken zich onderscheiden van de concurrentie en meer jonge sollicitanten aantrekken (Heineken, geciteerd in McCarthy et al., 2017). Zulk soort moderne selectiemethoden zijn voor organisaties een manier om te laten zien dat zij met de tijd meegaan en hebben, door gebruik te maken van het internet en sociale media, een groot bereik bij bepaalde doelgroepen. De efficiëntie en het gebruiksgemak van digitale selectiemethoden zijn voor veel organisaties een vanzelfsprekendheid geworden. Zo is de testcomputer niet meer weg te denken uit menig assessment center en kunnen we ons bijna niet meer voorstellen dat testresultaten tot een aantal jaar geleden met de hand gescoord moesten

worden. Dat organisaties veel baat hebben bij de digitalisatie van het selectieproces, blijkt wel uit de tijd- en kostenbesparing die zij daarmee doen (Anderson, 2003). Het vertrouwen van sommige HR professionals in digitale selectiemethoden is zelfs zo sterk, dat hun perceptie van bruikbaarheid en gebruiksvriendelijkheid een sterkere voorspeller vormt dan hun perceptie van voorspellende kracht en eerlijkheid voor het daadwerkelijk willen inzetten van technologische selectiemethoden (Oostrom, Van der Linden, Born, & Van der Molen, 2013). Door de winst in efficiëntie, het gebruiksgemak en het aantrekkelijk maken van het eigen merk lijkt het belang van technologie in het selectieproces voor organisaties ontegenzeggelijk.

Ook voor kandidaten is het belangrijk om een aantrekkelijke selectieprocedure te doorlopen. De perceptie van sollicitanten tijdens selectieprocedures heeft namelijk verstrekkende gevolgen op hun attitudes, intenties en gedragingen. Er zijn aanwijzingen dat zij bij een aantrekkelijke procedure sneller zullen solliciteren bij een organisatie en de werkgever vaker zullen aanbevelen bij kennissen en relaties (Hausknecht, Day & Thomas, 2004; McCarthy et al., 2017). Ook de motivatie en testprestatie zal toenemen wanneer sollicitanten de selectieprocedure als prettig en aangenaam ervaart (McCarthy et al., 2017). Kandidaten reageren bijvoorbeeld positiever op selectiemethoden zoals de *work sample* dan op cognitieve capaciteitentests en persoonlijkheidstests (Anderson, Salgado & Hülshager, 2010; Steiner & Gilliland, 1996). Onderzoek naar kandidaat percepties in het selectieproces heeft dan ook veel aandacht gekregen in de afgelopen decennia (Anderson et al., 2010; Gilliland, 1993; Hausknecht et al., 2004; McCarthy et al., 2017). Kandidaten vinden een selectiemethode of test veelal aantrekkelijk wanneer deze als eerlijk, transparant en relevant voor toekomstige werkzaamheden wordt ervaren (Anderson et al., 2010; Gililand, 1993). Gililand (1993) beschrijft hoe procedurele rechtvaardigheid, bekend uit de *organizational justice theory*, een belangrijke rol speelt bij de perceptievorming van de kandidaat tijdens het selectieproces. Selectiemethoden die gelijkenis tonen met de toekomstige functie (hoge procedurele rechtvaardigheid) worden als eerlijker gezien dan methoden die ogenschijnlijk niets met de functie te maken hebben, zoals cognitieve capaciteitentests. Kandidaten vinden selectiemethoden met een hoge procedurele rechtvaardigheid eerlijker, wat op zijn beurt zorgt voor onder andere een hogere testmotivatie en meer begrip en acceptatie voor het selectieproces.

Patterson, Zibarras, Carr, Irish & Gregory (2011) vinden dat selectiemethoden met een hoge waarheidsgetrouwheid (zoals simulaties) positiever en als eerlijker worden gewaardeerd

dan methoden met een lage waarheidsgetrouwheid (zoals een schriftelijke situationele beoordelingstest). Het probleem met selectiemethoden zoals simulaties en het rollenspel is echter dat ze relatief duur zijn en slechts bij kleine groepen kandidaten (meestal de laatst overgebleven kandidaten) worden ingezet (Lievens & De Soete, 2011). Daarnaast hebben rollenspellen een beperkte mate van standaardisatie. Het gedrag van de kandidaat is immers van invloed op het spel van de acteur en vice versa. Door het gebruik van multimedia en nieuwe technologie, zoals *virtual reality*, kan de waarheidsgetrouwheid van selectiemethoden en tests worden verhoogd. Audiovisuele stimuli kunnen worden gebruikt om tests te verrijken en van een hogere waarheidsgetrouwheid te voorzien, wat een positief effect kan hebben op testmotivatie- en prestatie en andere positieve kandidaat percepties in het selectieproces.

Dit scriptieonderzoek is gericht op de ontwikkeling van een situationele beoordelingstest in *virtual reality*. De test is ontwikkeld om inzicht in effectief leiderschap te meten. Voor zover bekend is dit de eerste keer dat *virtual reality* technologie wordt toegepast om een situationele beoordelingstest te construeren. Centraal staat dan ook de vraag of er voldoende psychometrische basis zal worden gevonden voor een situationele beoordelingstest in deze hoedanigheid. Daartoe wordt de vergelijking gemaakt met reeds bestaande tests en er zal worden gekeken naar de kandidaatpercepties van dit nieuw ontwikkelde instrument. De theoretische achtergrond van de situationele beoordelingstest en *virtual reality* technologie zullen hieronder worden toegelicht waarna de hypothesen van het huidige onderzoek worden ingeleid.

Situationele beoordelingstest

Een van de selectiemethoden die zich goed leent voor modernisering en digitalisatie is de situationele beoordelingstest of zogeheten *situational judgment test* (SJT). De SJT is een methode die een kandidaat een realistische werk gerelateerde situatie voorlegt, traditioneel op schrift gepresenteerd, waarna er een aantal mogelijke responsen wordt gegeven (Campion, Ployhart & MacKenzie, 2014; Lievens, Peeters & Schollaert, 2008; Whetzel & McDaniel, 2009). Het is vervolgens aan de kandidaat om de responsen te selecteren die hij of zij het meest en het minst waarschijnlijk zou uitvoeren (Motowidlo, Dunette & Carter, 1990). Een vertaald voorbeeld van een SJT item uit Motowidlo et al., 1990 is het volgende:

U en een collega van een andere afdeling zijn verantwoordelijk voor het coördineren van een gezamenlijk project. Uw collega voert overduidelijk zijn taken niet uit en neemt geen verantwoordelijkheid. In deze situatie zou u:

___Meest waarschijnlijk ___Minst waarschijnlijk

1. De situatie bespreken met uw manager en hem vragen uw collega's manager aan te spreken.
2. Uw collega eraan herinneren dat u zijn hulp nodig heeft omdat het project niet succesvol kan worden afgerond zonder inzet van jullie beiden.
3. Hem confronteren met het feit dat hij zijn deel van het werk niet uitvoert en hem vertellen dat u niet alles alleen zult doen. Wanneer hij zijn verantwoordelijkheid niet neemt, zult u de zaak met zijn manager bespreken.
4. Er achter proberen te komen waarom hij zijn deel niet uitvoert en hem uitleggen dat u daarom meer werk moet uitvoeren, wat het project niet ten goede komt.
5. Iemand anders van zijn afdeling met het project laten helpen.

SJTs zijn breed inzetbaar en worden gebruikt om verscheidene concepten te meten, zoals emotionele intelligentie (Sharma, Gangopadhyay, Austin & Mandal, 2013), integriteit (De Meijer, Born, Van Zielst & Van der Molen, 2010) en leiderschap (Oostrom, Born, Serlie & Van der Molen, 2012). SJTs zijn populair omdat ze organisaties in staat stellen om op relatief snelle en goedkope wijze een grote groep kandidaten gestandaardiseerde en werk gerelateerde tests te laten afnemen (Lievens et al., 2008). Ze worden in de praktijk veelvuldig ingezet omdat ze zo veelzijdig, voor kandidaten aantrekkelijk en valide zijn (Lievens & Motowidlo, 2016).

De SJT heeft sinds 1990 op veel aandacht kunnen rekenen van verschillende onderzoekers. Meta-analytische studies hebben aangetoond dat de SJT een valide methode is om baanprestatie te voorspellen (McDaniel, Hartman, Whetzel & Grub, 2007; McDaniel, Morgeson, Finnegan & Campion, 2001) met een gemiddelde predictieve validiteit van $r = .20$. Daarnaast laten SJTs incrementele validiteit zien boven cognitieve capaciteiten, persoonlijkheid en een composiet van cognitieve capaciteiten en persoonlijkheid, al zijn de gevonden incrementen niet erg groot – variërend van .01 tot .07 (McDaniel et al., 2007). De predictieve en incrementele validiteit wordt gevonden voor zowel SJTs met kennisgeoriënteerde “*should-do*” items (bijv. ‘Hoe effectief vindt u deze reactie?’) als voor SJTs met gedragsgeoriënteerde “*would-do*” items (bijv. ‘Wat zou u meest waarschijnlijk doen?’) (Lievens & De Soete, 2011). Dat onderscheid in responsinstructies is belangrijk, omdat blijkt dat kennisgeoriënteerde items een sterker beroep doen op cognitieve capaciteiten en dat gedragsgeoriënteerde items juist meer verband houden met persoonlijkheid (McDaniel et al., 2007). Zodoende speelt de responsinstructie van een SJT een belangrijke rol bij de constructvaliditeit van het instrument.

De constructvaliditeit van SJTs blijkt in de praktijk lastig te bepalen en is vaak punt van discussie geweest (Guenole, Chernyshenko & Weekly, 2017; Lievens, 2017; McDaniel, List & Kepes, 2016). Sommige onderzoekers beweren dat SJTs voornamelijk praktische intelligentie meten, ofwel het vermogen om met alledaagse situaties en omgevingen om te gaan (Stemler & Sternberg, 2006). Anderen zijn van mening dat SJTs vooral procedurele kennis meten (Motowidlo, Hooper & Jackson, 2006). Dit zou volgens Motowidlo en collega's (2006) komen door de zogeheten *implicit trait policies* van een kandidaat. Deze verklaring stelt dat de mate van een persoonlijkheidseigenschap van een kandidaat bepalend is voor de beoordeling van situaties die een beroep doen op die desbetreffende eigenschap. Een verdraagzame kandidaat zal bijvoorbeeld geloven dat de relatie tussen de persoonlijkheidsdimensie verdraagzaamheid en effectief gedrag sterker is dan een minder verdraagzame kandidaat. Volgens de auteurs verklaart deze theorie waarom SJTs en persoonlijkheid vaak significant correleren (Motowidlo et al., 2006). De matige constructvaliditeit van SJTs is dan ook een van de grootste kritiekpunten op het instrument. SJT items zijn vaak heterogeen op itemniveau, in de zin dat items op meerdere concepten een beroep doen, en laten daarom doorgaans correlaties zien met constructen die niet gerelateerd zouden moeten zijn. Een item kan bijvoorbeeld even sterk laden op cognitieve capaciteiten als op het persoonlijkheidskenmerk verdraagzaamheid. In voorkomende gevallen blijkt het daarom enorm lastig om interpreteerbare factoren te extraheren uit de itemscores. Daarnaast laten SJT schalen vaak geen discriminante validiteit zien (McDaniel et al., 2016). Ten gevolge van deze nadelen, beschikken SJTs vaak over lage interne consistentie betrouwbaarheidscoëfficiënten (Catano, Brochu & Lamerson, 2012). Sommige onderzoekers (Guneole et al., 2017; Lievens, 2017) pleitten daarom voor een meer construct gedreven opbouw van SJTs om de interpreteerbaarheid van de scores te verbeteren (voor een uitgebreide review en suggesties, zie Guenole et al., 2017). Daarentegen kan beargumenteerd worden dat het heterogene karakter van SJTs juist een voordeel is voor het voorspellen van werkprestatie (dat bestaat immers uit gedrag en handelingen die ook vaak heterogeen zijn).

Ondanks de verschillende opvattingen over de constructvaliditeit, zijn de meeste onderzoekers het eens dat de scores van een SJT meerdere soorten kennis en vaardigheden reflecteren, dat SJTs over relatief sterke criteriumvaliditeit beschikken en dat ze kleine tot gemiddeld-kleine subgroep verschillen laten zien (Campion et al., 2014). Dat laatste punt is interessant, omdat dit het instrument eerlijker maakt voor een grotere groep kandidaten. Uit

onderzoek blijkt dat vooral multimediale SJTs, die gebruik maken van technologie en multimedia, kleinere etnische verschillen laten zien (Chan & Schmitt, 1997; Lievens & Coetsier, 2002; Oostrom, De Soete & Lievens, 2015).

Waar de SJT van oorsprong schriftelijk werd afgenomen (Motowidlo et al., 1990) wordt er steeds meer gebruik gemaakt van multimediale tools om een SJT te ontwikkelen en af te nemen (Lievens & De Soete, 2011; Lievens & Sackett, 2006; Oostrom et al., 2015; Whetzel & McDaniel, 2009). De video-SJT, waar de situaties door acteurs in een video worden gepresenteerd, is inmiddels een veelgebruikte en breed geaccepteerde vorm van *situational judgment testing* (Lievens & Sackett, 2006; Olson-Buchanan & Drasgow, 2006). Dit soort multimedia SJTs heeft veel aandacht gekregen in het onderzoeksveld en laat zien onder andere baanprestatie (Lievens & Sackett, 2006) en academisch succes (Oostrom, Born, Serlie & Van der Molen, 2011) te voorspellen. Door de grote gelijkenissen met echte werkzaamheden en de audiovisuele presentatie van de situaties, wordt bij dit soort SJTs in de literatuur gesproken over simulaties met een hoge waarheidsgetrouwheid. Op tekst gebaseerde SJTs hebben dan dus ook een lagere waarheidsgetrouwheid. Op het eerste gezicht lijken de voordelen van deze simulaties met een hoge waarheidsgetrouwheid vanzelfsprekend. De werksituaties worden immers meer realistisch weergegeven doordat er gebruik gemaakt kan worden van verschillende verbale en non-verbale cues, zoals de gezichtsuitdrukking en de intonatie van een acteur (Olson-Buchanan & Drasgow, 2006).

Het gebruik van multimediale SJTs laat een aantal veelbelovende voordelen zien ten opzichte van de traditionele, op schrift gepresenteerde SJTs. Zo reageren kandidaten er positiever op (Chan & Schmitt, 1997; Kanning, Grew, Hollenberg & Hadouch, 2006; Richman-Hirsch, Olson-Buchanan & Drasgow, 2000), wordt er een hogere criteriumvaliditeit gevonden dan bij schriftelijke SJTs (Christian, Edwards & Bradley, 2010), is de predictieve validiteit voor interpersoonlijke vaardigheden hoger (Lievens & Sackett, 2006) en laten multimedia SJTs minder gender en etnische verschillen zien dan traditionele SJTs (Chan & Schmitt, 1997; Lievens & Coetsier, 2002). Dat laatste komt waarschijnlijk doordat multimedia SJTs minder een beroep doen op cognitieve vaardigheden dan tekst-gebaseerde tests (Lievens et al., 2008; Lievens & Sackett, 2006).

Naast de vele voordelen van multimedia SJTs, worden er in de literatuur ook een aantal nadelen genoemd. Ten eerste vergt de ontwikkeling van een multimedia SJT meer tijd en geld

dan de schriftelijke variant. Daarnaast bestaat de mogelijkheid dat multimedia SJTs irrelevante contextuele informatie toevoegen, zoals het uiterlijk van een acteur, waardoor de kans op error groter wordt (Whetzel & McDaniel, 2009). De al dan niet negatieve effecten van dit soort contextuele informatie zijn niet in zicht, omdat dit in de literatuur nauwelijks wordt besproken. Over de gehele lijn bieden Multimedia SJTs dus vele voordelen ten opzichte van op tekst gebaseerde SJTs. Hoewel deze SJTs realistischer zijn dan tekstuele SJTs, worden situaties en/of reacties nog altijd getoond op een vlak beeldscherm. Deelnemers zullen zich daarom in beperkte mate onderdeel voelen van de situatie. Ook heeft de deelnemer weinig vrijheid om zelf te bepalen waarop de aandacht wordt gericht. Deze wordt immers sterk gestuurd door de opstelling van de camera en de montage van de regisseur. Daarnaast blijft een minpunt van alle soorten SJTs dat de geschetste situaties kort (één situatie of gebeurtenis per item) en dus niet erg meeslepend zijn. Ook dit kan ten koste van het realisme, de diepte van de geschetste situatie en dus de waarheidsgetrouwheid gaan.

Een doorontwikkeling van multimedia SJTs die deze problemen oplost, is de *virtual reality* (VR) SJT. In een virtuele omgeving kan de kandidaat immers zelf bepalen waar de aandacht op wordt gericht en kunnen werksituaties levensecht worden weergegeven en worden ervaren. De toepassing van deze technologie zorgt voor een waarheidsgetrouwheid die simulaties en rollenspellen benadert.

Virtual reality

In essentie is VR de computergestuurde techniek om mensen in een alternatieve, virtuele wereld onder te dompelen. Doorgaans wordt dit bewerkstelligd door het gebruik van een speciale VR-bril en koptelefoon, die ervoor zorgen dat iemand een virtuele wereld kan waarnemen en tegelijkertijd wordt afgesloten van audiovisuele prikkels van de fysieke wereld. Dit geeft mensen het gevoel zich daadwerkelijk in die virtuele wereld te bevinden en ontlokt een gevoel van *presence* (Hartmann et al., 2016). Dit gevoel van *presence* is een psychologische staat waar VR gebruikers het idee krijgen zich daadwerkelijk te bevinden in en te interacteren met een virtuele omgeving (van Gelder et al., 2018). In een recentelijk onderzoek werd een emotioneel beladen situatie (een gevecht) schriftelijk en in VR gepresenteerd aan twee groepen deelnemers (Van Gelder, de Vries, Demetriou, van Sintemaartensdijk & Donker, 2019). De auteurs toonden aan dat kandidaten die een situatie in VR beleven, meer realisme en *presence* ervaren dan kandidaten die de schriftelijke situatie tot zich nemen en er bleek verder dat de VR situatie meer emotionele

responsen ontlokte (van Gelder et al., 2019). VR-technologie lijkt daarmee een uitgelezen manier om tests en SJTs van meer realisme en waarheidsgetrouwheid te voorzien.

Aguinis, Henle & Beaty (2001) rapporteerden bijna twintig jaar geleden al over de mogelijke toepassingen van VR binnen het selectieproces. Zij beschrijven hoe VR gebruikt kan worden om verschillende vaardigheden te meten, zoals een oefening om met virtuele ondergeschikten te interacteren als test voor interpersoonlijke en leidinggevende vaardigheden. Ook zien zij mogelijkheden om persoonlijkheidskarakteristieken zoals consciëntieusheid en integriteit te toetsen in VR, door een kandidaat in een virtuele kantooromgeving taken te laten uitvoeren. Als potentiële voordelen noemen zij de hoge waarheidsgetrouwheid, gevoelens van *presence* en hoge indrukvaliditeit, het feit dat standaardisatie gewaarborgd kan worden in tegenstelling tot andere simulaties en rollenspellen en het feit dat er kleinere etnische subgroep verschillen kunnen worden gevonden (Aguinis et al., 2001).

Interessant om te zien is de prijslijst van VR materialen die Aguinis en collega's (2001) noemen in hun artikel. Een degelijke VR setup kostte destijds tienduizenden dollars, waar er bovendien prijzige softwarelicenties dienden te worden afgesloten. Hierdoor beperkte het gebruik van VR zich tot enkele vermogende organisaties, zoals het Amerikaanse defensieapparaat en NASA. Door recentelijke technologische ontwikkelingen is VR veel betaalbaarder en toegankelijker geworden. Tegenwoordig is de gemiddelde smartphone krachtig genoeg om VR applicaties te draaien en er zijn VR brillen verkrijgbaar vanaf slechts tien euro (Bailenson, 2018). Deze ontwikkelingen maken VR veel toegankelijker voor een breder publiek en zorgen ervoor dat VR binnen de meeste onderzoeksbudgetten zal passen.

In de laatste jaren wordt er in het klinisch psychologisch onderzoeksveld veel gebruikt gemaakt van VR-technologie, vooral bij de behandeling van angststoornissen (Diemer, Alpers, Peperkorn, Shiban & Mühlberger, 2015; Freeman et al., 2017). VR blijkt effectief bij het behandelen van sociale angst (Anderson et al., 2013), posttraumatische stress (Rothbaum et al., 2014) en wordt gebruikt om sociale vaardigheden aan te leren bij jongvolwassenen met autisme (Kandalaf, Didehbani, Krawczyk, Allen & Chapman, 2013). Opmerkelijk genoeg wordt er in het veld van de arbeids- en organisatiepsychologie vrijwel geen onderzoek gedaan naar de toepassingen van VR, ondanks de grote technologische ontwikkelingen en de aandacht voor (andere) technologie bij het selecteren en trainen van personeel (Lievens & De Soete, 2011; Ostrom et al., 2013).

Huidige onderzoek

In het huidige onderzoek wordt gerapporteerd over de ontwikkeling en in het bijzonder de psychometrische eigenschappen van een nieuw ontwikkelde VR-SJT. In deze VR-SJT wordt de kandidaat meegenomen naar een virtuele vergaderomgeving waar een team van vier medewerkers en hun leidinggevende een vergadering voeren, die van aanvang tot afsluiting zal duren. Figuren 1 en 2 geven een indruk van de omgeving. De kandidaat heeft de rol van observator en observeert de verschillende rollen, karakters en belangen van de teamleden. De VR-SJT beoogt het inzicht in effectief leiderschap te meten. De kandidaat zal daarom gevraagd worden het gedrag en het optreden van de leidinggevende te beoordelen op effectiviteit. Anders dan bij bestaande SJTs, houdt de VR-SJT zich bij één setting (de vergaderruimte) waarin zich verschillende gebeurtenissen voordoen. Bestaande SJTs bevatten voornamelijk een reeks van korte, onafhankelijke situaties en omgevingen die beoordeeld dienen te worden (Lievens et al., 2008). Het is dus goed mogelijk dat participanten zich beter kunnen inleven in – en verbonden zullen voelen met – de personages in de VR-SJT. VR-technologie biedt talrijke nieuwe mogelijkheden die bij op tekst gebaseerde en reeds bestaande multimedia SJTs niet mogelijk zijn. Kandidaten worden in de VR-SJT volledig ondergedompeld in de virtuele wereld, waardoor zij niet afgeleid kunnen raken door zaken van buitenaf. Hierdoor zullen kandidaten dus meer gefocust zijn op de test en kunnen ze hun volledige aandacht schenken aan het doorlopen van de VR-SJT. Meer aandacht voor een test heeft dan ook een positief effect op de testprestatie van een kandidaat (McCarthy et al., 2017).

Daarnaast kan de kandidaat zelf bepalen waarop en op wie hij zijn aandacht richt, door met de virtualrealitybril in het rond te kijken. Op die manier kunnen er ook groepsprocessen worden waargenomen. Er kan bijvoorbeeld worden waargenomen hoe het gedrag van een leidinggevende het team beïnvloedt, wat de kandidaat zal helpen in het beoordelen van de effectiviteit van de leidinggevende. Een belangrijk voordeel van de VR-SJT is dat werkgerelateerde situaties veel realistischer kunnen worden weergegeven dan bij ieder andere soort (multimedia) SJT. De kandidaat zal de situaties daarom veel meer gaan beleven en ervaren dan deze slechts te aanschouwen, wat de waarheidsgetrouwheid van het instrument ten goede zal komen.

Vanwege verschillende, meestal financiële, redenen vindt de interactie in simulaties en multimedia SJTs zich vrijwel altijd plaats in 1-op-1 gesprekken. Het is echter interessant om de

interactie in teams waar te nemen omdat dat voor de meeste mensen een belangrijk onderdeel vormt voor werkgedrag in organisaties. Het vergaderen met teamleden vormt bijvoorbeeld een van de meest kenmerkende en toegepaste vormen van leidinggeven (Odermatt et al., 2017). Odermatt en collega's (2017) vinden dat het gedrag van een leidinggevende een belangrijke rol speelt bij het verloop van een vergadering. In het specifiek vinden zij dat persoonsgericht leidinggevend gedrag een wezenlijke rol speelt bij het positief leiden van een vergadering. Persoonsgericht leidinggevend hebben meer aandacht voor open en respectvolle communicatie en letten meer op de persoonlijke behoeften van hun teamleden, zoals betrokkenheid en acceptatie binnen de vergadering. Teamleden die onder een persoonsgericht leidinggevende vergaderden, waren daarom meer tevreden over de vergadering. Opmerkelijk genoeg vinden Odermatt en collega's (2017) geen verband tussen taakgericht leidinggevend gedrag en de effectiviteit van de vergadering. Taakgericht leidinggevend bieden structuur aan de vergadering en zijn meer doelgericht in hun aanpak. Een mogelijke verklaring wordt gegeven in het feit dat vergaderingen van nature al gestructureerd zijn, waardoor het belang van een leidinggevende die structuur aanbrengt wegvalt. Het belang van een goede leider in vergaderomgevingen wordt des te meer duidelijk in een ander onderzoek van Odermatt en collega's (2018). In dit onderzoek vonden zij dat werknemers die tijdens een vergadering zogeheten *uncivil meeting behaviors* (UMBs) lieten zien, een negatief effect hadden op de tevredenheid van andere werknemers en de effectiviteit van de vergadering. UMBs worden beschreven als gedragingen die blijk geven van onbeleefdheid, ongepastheid en desinteresse (Oldermatt et al., 2018). In voorkomende gevallen is het dus belangrijk dat een competente leidinggevende de vergadering leidt, om dit soort negatieve gedragingen te beperken en/of voorkomen.

Voor (toekomstig) leidinggevend is het belangrijk om kennis van, en inzicht in, eerdergenoemde leiderschapsstijlen te hebben. De VR-SJT is ontworpen om inzicht in effectief leiderschap te meten. (Inzicht in) leiderschap en SJTs blijken een goede match; uit een meta-analyse van Christian en collega's (2010) blijkt dan ook dat de meeste SJTs dit domein toetsen en dat dit soort SJTs de sterkste predictieve validiteit voor werkprestatie laten zien. De VR-SJT doet een beroep op de eerdergenoemde constructen taakgericht en persoonsgericht leidinggeven. Deze constructen zijn gebaseerd op de dimensies *Initiating structure* en *Consideration* waar veel leiderschapsgedrag naar is terug te leiden (Derue, Nahrgang, Wellman, & Humphrey, 2011;

Judge, Piccolo & Ilies, 2004). Het soort leidinggevend gedrag is dan ook een belangrijke voorspeller voor effectief leiderschap. Zo is meer taakgericht (*initiating structure*) leidinggevend gedrag (zoals het directief aanspreken van personeel) voorspellend voor prestaties en is meer persoonsgericht (*consideration*) leidinggevend gedrag (zoals het ondersteunen en laten ontwikkelen van personeel) voorspellend voor de tevredenheid met de leidinggevende (Derue et al., 2011). De effectiviteit en de geschiktheid van beide soort gedragingen hangt natuurlijk af van de situatie die zich voordoet. In bepaalde gevallen zal een meer directieve stijl van leidinggeven effectiever zijn dan een ondersteunende stijl en vice versa. Een competente leider zal per definitie aanvoelen welk soort gedrag het meest geschikt is voor welke situatie. Dit doet hij of zij op basis van ervaring, kennis, training en persoonlijkheid; constructen die in verschillende mate ook door SJTs worden gemeten (Guenole et al., 2017). Tijdens de VR-SJT zal de kandidaat dan ook verschillende situaties te zien krijgen waarin beide stijlen van leidinggeven effectief en minder effectief worden toegepast. Het is vervolgens aan de kandidaat om de effectiviteit van de leidinggevende te evalueren. Zie de methode voor uitleg en een voorbeeld van de items.



Figuur 1. Linkerkant van de vergaderomgeving zoals kandidaten die te zien krijgen.



Figuur 2. Rechterkant van de vergaderomgeving zoals kandidaten die te zien krijgen, met rechts Chris Moorman, de leidinggevende van het team.

Hypothesen

Centraal staat de vraag of er voldoende psychometrische basis zal worden gevonden voor een SJT in VR format.

In het bijzonder zal de vergelijking worden gemaakt met de VR-SJT en de Sociaal quotiënt test Leidinggeven (SQ-leid). De SQ-leid is een multimediale (video) SJT met een vergelijkbare meetpretentie. Deze video-SJT meet het sociaal inzicht in leidinggevende situaties en rapporteert naast een totaalscore ook scores op de vier domeinen aanspreken op resultaten, aanspreken op (sociaal) gedrag, motiveren en coachen. Onderzoek naar de predictieve validiteit van de SQ-leid laat zien dat deze video-SJT academisch succes voorspelt, met een geobserveerde correlatie van .28 (Op de Beek, Oostrom & Born, 2011).

Het aanspreken op resultaten en gedrag zijn vormen van taakgericht leidinggeven en het motiveren en coachen vallen onder het persoonsgericht leidinggeven. Zodoende zal de soortgenootvaliditeit van de VR-SJT en de video-SJT worden onderzocht. Door dezelfde soort meetpretentie wordt er een gemiddelde correlatie verwacht tussen beide SJTs.

Hypothese 1: De VR-SJT doet een beroep op de constructen taakgericht leidinggeven en persoonsgericht leidinggeven.

Hypothese 2: De VR-SJT heeft een positieve gemiddelde correlatie met de video-SJT ($.30 \leq r \leq .50$).

Effectief leiderschap is situatieafhankelijk en is van nature een heterogeen concept waarin een beroep wordt gedaan op zowel taakgericht als persoonsgericht leidinggeven (Derue et al., 2011). Daarom wordt er dan ook verwacht dat de constructen taakgericht leidinggeven en persoonsgericht leidinggeven positief zullen correleren.

Hypothese 3: De constructen taakgericht leidinggeven en persoonsgericht leidinggeven in de VR-SJT zijn positief gemiddeld met elkaar gecorreleerd ($.30 \leq r \leq .50$).

Zoals eerder vermeld, is de betrouwbaarheid van SJTs doorgaans lastig in kaart te brengen, omdat ze vaak geen homogene concepten meten (Catano et al., 2012; Guenole et al., 2017; McDaniel et al., 2016; Sorrel et al., 2016; Whetzel & McDaniel, 2009). De interne consistentie van de test kan in dat geval een onderschatting zijn van de betrouwbaarheid (Catano et al., 2012; Lievens et al., 2008). Catano en collega's (2012) vinden bijvoorbeeld een alfacoëfficiënt van 'slechts' $\alpha = .46$ na een meta-analyse van 56 alfacoëfficiënten vanuit de SJT literatuur. Ook de gestratificeerde alfacoëfficiënten wijken nauwelijks af van de reguliere Cronbach's alfacoëfficiënten die zij in hun onderzoek vinden (Catano et al., 2012). De test-hertest betrouwbaarheid kan in dat geval als alternatief worden gebruikt (Catano et al., 2012; Lievens et al., 2008). Het valideren van SJTs door middel van de test-hertest betrouwbaarheid blijkt in de praktijk vaak niet realiseerbaar omdat financiële en tijdgebonden factoren hierin een beperkende rol vormen.

Door deze praktische beperkingen kan er ook in het huidige onderzoek geen gebruik worden gemaakt van de test-hertest betrouwbaarheid. Ondanks de aanbevelingen om de test-hertest betrouwbaarheid te gebruiken bij SJTs, zal er in het huidige onderzoek toch gebruik worden gemaakt van de interne consistentie om de betrouwbaarheid van het instrument te toetsen. Omdat de VR-SJT een beroep doet op twee constructen, zullen er twee interne consistentie coëfficiënten worden gemeten. Daarnaast zal ook de alfacoëfficiënt voor de totaalscore worden berekend.

Hypothese 4a: De VR-SJT beschikt over voldoende betrouwbaarheid ($\alpha \geq .70$) voor de totaalscore.

Hypothese 4b: De VR-SJT beschikt over voldoende betrouwbaarheid ($\alpha \geq .70$) voor

het construct taakgericht leidinggeven.

Hypothese 4c: De VR-SJT beschikt over voldoende betrouwbaarheid ($\alpha \geq .70$) voor het construct persoonsgericht leidinggeven.

Omdat de VR-SJT vooral uit op kennisgeoriënteerde items bestaat, wordt er een correlatie met cognitieve capaciteiten verwacht (McDaniel et al., 2007; Oostrom et al., 2015). Deze zogeheten “*should-do*” responsinstructie laat namelijk een sterkere relatie met cognitief vermogen zien (McDaniel et al., 2007). Logischerwijs volgt dat de VR-SJT dus ook minder gedragsgeoriënteerde, of zogeheten “*would-do*” items bevat. Dit soort vragen doen een beroep op gedragstendensen en vertonen een sterkere relatie met persoonlijkheid (McDaniel et al., 2007; Oostrom et al., 2015). Zodoende wordt er in het huidige onderzoek een lagere correlatie verwacht met persoonlijkheidskenmerken.

Hypothese 5: De VR-SJT toont een hogere correlatiecoëfficiënt met cognitieve capaciteiten dan met persoonlijkheidskenmerken.

Door het toegevoegde realisme, de hoge waarheidsgetrouwheid en de ‘onderdompelende’ ervaring die virtual reality biedt, wordt er tot slot verwacht dat kandidaten de VR-SJT als aantrekkelijker ervaren en positiever beoordelen dan een video-SJT (SQ-leid).

Hypothese 6: De VR-SJT wordt positiever beoordeeld dan de video-SJT.

Methode

Deelnemers en procedure

In het huidige onderzoek deden 53 vrouwen en 48 mannen mee ($N = 101$). De gemiddelde leeftijd van de steekproef was 43 jaar ($M = 43.04$, $SD = 10.96$). Hoger opgeleiden waren sterk vertegenwoordigd in de steekproef; 46 participanten zijn hbo-opgeleid en 40 participanten zijn wo-opgeleid. Slechts 4 participanten zijn mbo-opgeleid, 1 participant is vwo-opgeleid en van 10 participanten is geen opleidingsniveau bekend. De steekproef van dit onderzoek bestaat uit drie groepen deelnemers, die hieronder verder worden beschreven.

Experts. In totaal zijn 28 *subject matter experts* (SMEs) getest (15 vrouwen, 13 mannen) op hun inzicht in effectief leiderschap. De scores van de SMEs zijn bepalend voor de normering

van de VR-SJT. SMEs zijn vanuit professionele netwerken benaderd met de vraag mee te willen werken aan het huidige onderzoek ten behoeve van de normering van dit nieuwe instrument. De SMEs zijn geselecteerd op een aantal belangrijke criteria. Zij moeten voldoende leidinggevende ervaring hebben opgedaan, moeten op de hoogte zijn van de eisen die aan medewerkers en ondergeschikten worden gesteld en moeten goed zicht hebben op de kwaliteiten van medewerkers in hun betreffende organisatie. Zo kan verondersteld worden dat zij snappen welk leidinggevend gedrag effectief is en welk gedrag dat niet is in een teamvergadering. Daarnaast komen ook assessment- en arbeidspsychologen in aanmerking om op te treden als SME. Zij beschikken immers over de meta kennis van het leidinggeven en zouden dus goed in staat moeten zijn om effectief en ineffectief leidinggevend gedrag te herkennen. De daadwerkelijke selectie van SMEs is gemaakt door een drietal partijen, te noemen Van der Maesen | Koch HRM-advies, SLIM assessments en eelloo. Deze partijen zijn betrokken bij de ontwikkeling van de VR-SJT. Van de 32 benaderde SMEs hebben er 28 de VR-SJT doorlopen, wat leidt tot een responspercentage van 87.5%.

Assessmentkandidaten. Verder zijn er 57 assessmentkandidaten (32 vrouwen, 25 mannen) getest ($M = 44.82$ jaar, $SD = 9.46$ jaar), die beperkte ervaring hebben met leidinggeven of die leidinggevende ambities hebben. De scores van de assessmentkandidaten worden gebruikt ter validatie van de VR-SJT. Kandidatendata is verzameld vanuit zowel selectie als ontwikkel- en loopbaanassessments. De kandidaten die hebben deelgenomen aan de VR-SJT deden een assessment voor een toekomstig leidinggevende functie. De daadwerkelijke selectie van deelnemende assessmentkandidaten is gemaakt door SLIM Assessments, waar de VR-SJT tijdens de assessmentdag op locatie is afgenomen. Van de 57 geselecteerde assessmentkandidaten hebben er 57 de VR-SJT doorlopen, wat leidt tot een responspercentage van 100%.

Persoonlijke relaties. Tot slot zijn er 16 participanten (6 vrouwen, 10 mannen) getest ($M = 34.29$ jaar, $SD = 11.85$ jaar) vanuit de privésfeer om bij te dragen aan de validatie van het instrument. Bij deze participanten is er geen voorselectie gemaakt op basis van leidinggevende ervaring en/of leidinggevende ambities. Van de 21 benaderde participanten hebben er 16 de VR-SJT doorlopen, wat leidt tot een responspercentage van 76.2%.

Materialen

Inzicht in effectief leiderschap.

Virtual reality situational judgment test (VR-SJT). Centraal staat het nieuw ontwikkelde instrument, de VR-SJT. Tijdens deze test wordt de kandidaat meegenomen naar een virtuele vergaderomgeving waar een team van vier medewerkers en hun leidinggevende een vergadering voeren, die van aanvang tot afsluiting zal duren. Deze test is ontwikkeld door Van der Maesen | Koch HRM-advies in samenwerking met SLIM assessments en eelloo. De VR-SJT is gefilmd en bewerkt door een productiebedrijf en gebruikt professionele acteurs. Het instrument bestaat uit een 360° video met 11 scènes die ieder afgewisseld worden door een reeks van ca. 8 vragen. De kandidaat beoordeelt in totaal 84 items, $\alpha = .87$. De VR-SJT begint met een voorbeeldsituatie en een reeks voorbeeldvragen, om de kandidaat te laten wennen aan de ervaring van VR en aan de test. De netto speeltijd van de scènes is 15 minuten, de duur van de vragenreeks hangt af van het invulgedrag van de kandidaat. De volledige tijdsduur van de VR-SJT is ongeveer drie kwartier ($M = 49$ minuten, $SD = 13$ minuten).

De VR-SJT wordt afgenomen middels een Samsung Gear VR SM-325 virtualrealitybril. Hiermee kan de deelnemer 360 graden rond zich heen kijken. Een stereo koptelefoon (JBL T450) voorziet de deelnemer tevens van 3D audio tijdens de test. De daadwerkelijke test wordt afgespeeld via de 'VR-assessment' app die is geïnstalleerd op een Samsung S8 smartphone. Deze applicatie is specifiek voor dit doeleinde ontwikkeld en communiceert met een online testplatform waar de scores worden opgeslagen.

De ontwikkeling van de VR-SJT heeft plaatsgevonden volgens stappen die gebruikelijk zijn voor het ontwikkelen van (multimedia) SJTs (Lievens et al., 2008). Hieronder worden enkele stappen nader beschreven. Allereerst zijn kritieke incidenten verzameld die de basis vormen voor de item content en dus de geschetste situaties. Dit zijn situaties waar leidinggevendend regelmatig mee te maken krijgen op hun werk. De kritieke incidenten zijn verzameld bij ervaren leidinggevendend en SMEs en hebben betrekking op taak- en persoonsgericht leidinggevendend. De kritieke incidenten zijn vervolgens herschreven naar een script en verfilmd. Een voorbeeld van een kritiek incident is een medewerker die zich niet constructief opstelt en weerstand vertoont tijdens de vergadering. Een tweede belangrijke fase bij het ontwikkelen van SJTs, is het creëren van de responsen. De responsen zijn in dit geval het al dan niet effectieve leidinggevendend gedrag van de acteur tijdens een scène. Bij het creëren van deze responsen zijn eveneens SMEs geraadpleegd. Een voorbeeld van een respons is de leidinggevendend die eerdergenoemde medewerker aanspreekt op zijn gebrek aan participatie.

Anders dan bij de meeste SJTs, wordt er in de VR-SJT slechts een enkele respons op een situatie behandeld. De acteur treedt immers maar een keer op (respons) bij een bepaalde gebeurtenis (kritiek incident). Dit soort SJTs worden ook wel *single-respons* SJTs genoemd en laten ook voldoende voorspellende waarde zien voor werkprestatie (Crook et al., 2011; Motowidlo, Crook, Kell & Naemi, 2009).

Zoals eerder vermeld, worden er 84 responsen (items) getoond tijdens de VR-SJT. De deelnemer beoordeelde de effectiviteit van iedere van deze responsen op een 5-punts Likertschaal. Er is bijvoorbeeld een situatie waarin een van de teamleden te laat de vergaderruimte binnenkomt. De leidinggevende spreekt hem vervolgens op vrij directieve wijze aan op zijn late binnenkomst. Een item over deze situatie luidt als volgt:

Chris gaat op effectieve wijze om met de late binnenkomst van Daan.

Wat is hierover uw oordeel?

Helemaal oneens | -- - +/- + ++ | Helemaal eens

Zodoende wordt er uiteindelijk duidelijk hoeveel inzicht de kandidaat heeft in effectief leiderschapsgedrag. De beoordeling van de kandidaat wordt afgezet tegen de gemiddelde beoordeling van 28 SMEs (met een interbeoordelaarsbetrouwbaarheid van $\alpha = .94$), zodat hoe hoger de overeenstemming tussen de kandidaat en de SMEs, hoe hoger de score van de kandidaat op de VR-SJT. De VR-SJT rapporteert ten slotte naast een totaalscore op de schaal taakgericht leidinggeven, $\alpha = .71$ en op de schaal persoonsgericht leidinggeven, $\alpha = .73$.

Video-SJT (SQ-leid). De SQ-leid is een bestaande en reeds gevalideerde video-SJT (Van der Maesen | Koch HRM-advies, 2011a). De SQ-leid is ontwikkeld om sociaal inzicht in leidinggevend gedrag te meten en bestaat uit 17 werksituaties die leidinggevend regelmatig kunnen tegenkomen op hun werk. De situaties worden door een verteller geïntroduceerd. Na elke situatieschets krijgt de kandidaat 4 korte responsen te zien die per video worden aangeboden. De responsen verschillen in oplossing, gedrag en stijl van leidinggeven. De kandidaat beoordeelde elk van deze videofragmenten op effectiviteit middels een 5-punts Likertschaal die varieert van (--) *helemaal niet effectief* tot (++) *heel effectief*. In totaal worden er 68 items beoordeeld, $\alpha = .82$. De beoordeling van de kandidaat wordt afgezet tegen de gemiddelde beoordeling van 15 SMEs (met een interbeoordelaarsbetrouwbaarheid van $\alpha = .93$), zodat hoe hoger de overeenstemming tussen de kandidaat en de SMEs, hoe hoger de score van de kandidaat op de

test. De SQ-leid rapporteert ten slotte op de volgende domeinen: het aanspreken op resultaten, het aanspreken op (sociaal) gedrag, het motiveren van medewerkers en het coachen van medewerkers. De domeinen aanspreken op resultaten en aanspreken op (sociaal) gedrag zijn in het huidige onderzoek samengenomen tot de schaal taakgericht leidinggeven, $\alpha = .72$. De domeinen motiveren en coachen zijn samengenomen tot de schaal persoonsgericht leidinggeven, $\alpha = .70$.

Cognitieve capaciteiten.

Internet cognitieve capaciteiten test (ICCT-HV). De Internet cognitieve capaciteiten test hoog niveau verlengde versie (ICCT-HV) is een cognitieve capaciteitentest die bestaat uit zes onderdelen: (1) analogieën, (2) diagrammen, (3) cijferreeksen, (4) figuurreeksen, (5) schatten, en (6) figuurklassen (Van der Maesen | Koch HRM-advies, 2011b). De onderdelen analogieën en diagrammen behoren tot het domeinen verbale intelligentie (35 items, $\alpha = .74$). De onderdelen cijferreeksen en schatten behoren tot het domein numerieke intelligentie (35 items, $\alpha = .80$). De onderdelen figuurklassen en figuurreeksen behoren tot het domein ruimtelijke intelligentie (30 items, $\alpha = .75$). In totaal bestaat de ICCT-HV uit 100 items, $\alpha = .85$. De alfacoefficiënten zijn gebaseerd op een normgroep uit 2014, $N = 1731$; niveau: 13% mbo, 60% hbo, 27% wo).

Persoonlijkheid.

HEXACO persoonlijkheidsvragenlijst. De Nederlandse vertaling van de HEXACO persoonlijkheidsvragenlijst gereviseerd is gebruikt om het persoonlijkheidsprofiel te meten (De Vries, Ashton, & Lee, 2009). De HEXACO bestaat uit 192 items en meet zes persoonlijkheidsdimensies, te weten integriteit (32 items, $\alpha = .91$), emotionaliteit (32 items, $\alpha = .88$), extraversie (32 items, $\alpha = .90$), verdraagzaamheid (32 items, $\alpha = .88$), consciëntieusheid (32 items, $\alpha = .85$) en openheid voor ervaringen (32 items, $\alpha = .87$). De interne consistentie is gebaseerd op een steekproef van de Nederlandse populatie, $N = 1352$ (De Vries et al., 2009).

Kandidaatperceptie. Tot slot is de kandidaatperceptie gemeten met een schaal van 6 items die voor dit onderzoek is ontwikkeld. De alfacoefficiënt varieerde per afnamemoment ($\alpha = .66 - .70$). Deze vragenlijst meet hoe positief de kandidaten een instrument beoordelen en bestaat uit de volgende items: “Ik vond deze test leuk om te doen”, “Het was me duidelijk wat er in deze test van mij werd verwacht”, “De inhoud van de test was realistisch en herkenbaar”, “Deze test komt op mij over als nieuw en modern”, “Deze test lijkt mij relevant voor het voorspellen van leidinggevende vaardigheden” en “Ik heb mijn best gedaan om deze test goed te maken”. De

items werden beoordeeld op een 5-punts Likertschaal die varieert van (--) *helemaal niet mee eens* tot (++) *helemaal mee eens*.

Resultaten

Testontwikkeling

Allereerst zijn enkele psychometrische eigenschappen van de VR-SJT onderzocht. Een betrouwbaarheidsanalyse van alle 84 VR-SJT items toonde dat een viertal items een negatieve item-totaal correlatie lieten zien. Het verwijderen van de desbetreffende items had een positief effect op de waarde van de betrouwbaarheidscoëfficiënt van de totaalscore en de subscores. De vier afwijkende items waren de volgende: “Chris verzekert zich er voldoende van dat haar mail is gelezen. Wat is hierover uw oordeel?”, “Als niemand met ideeën komt over de uitvoering zal Chris zelf met suggesties moeten komen. Wat is hierover uw oordeel?”, “Chris zou meer actie moeten ondernemen om de overleg sfeer beïnvloeden. Wat is hierover uw oordeel?” en “Chris benut de koffiepauze om Frank onder vier ogen te spreken. Vindt u dit een effectieve actie?”. Deze items waren problematisch omdat ze te weinig onderscheidend vermogen toonden en er erg weinig spreiding in de antwoorden zat. Het lijkt er daarom op dat deze items te gemakkelijk waren en dus van weinig toegevoegde waarde zijn voor de test. Op basis van bovengenoemde punten en de negatieve item-totaal correlaties zijn deze vier items verwijderd uit de test. Vervolgens zijn de scores van de SMEs ($n = 28$) nader bestudeerd. Een betrouwbaarheidsanalyse van de SME scores toonde dat de mate van overeenstemming tussen de experts hoog was; intra-klasse correlatiecoëfficiënt (ICC) bleek gelijk aan .94. Het scorepatroon van een van de experts week echter sterk af van de rest van de groep. De item-totaal correlatie van deze expert was laag ($r = .15$), waardoor deze deelnemer niet langer als SME werd opgenomen in de verdere analyses. De interbeoordelaarsbetrouwbaarheid bleef ongewijzigd na het verwijderen van de afwijkende deelnemer uit de expertgroep.

Na bovenstaande analyses en het vaststellen van de definitieve expertgroep zijn de ruwe kandidatenscores omgezet in normscores. De Leng en collega's (2017) beschrijven 28 manieren om SJTs te scoren en demonstreren de invloed van de scoring op de betrouwbaarheid van SJTs. Een van hun methoden, die gebruikmaakt van de ruwe consensus tussen de kandidaatscore en de gemiddelde expertscores, levert de hoogste betrouwbaarheidscoëfficiënt op. Deze methode is dan ook gebruikt om de kandidatendata te scoren in het huidige onderzoek. Voor ieder item werd de

absolute afstand bepaald tussen de beoordeling door de deelnemer en die door de expert. Het aantal verdiende punten per item werd bepaald door de verschilscore af te trekken van 5. De items worden immers beantwoord op een 5-puntsschaal. Bijvoorbeeld, een kandidaat beoordeelt een item met een 4 terwijl het expertgemiddelde 3.6 is. De absolute afstand is in dat geval 0.4. De kandidaat behaalt vervolgens een score van $5 - 0.4 = 4.6$.

Een kandidaat die verder van het expertoordeel verwijderd is, door bijvoorbeeld een 1 toe te kennen aan het item, heeft een absolute afstand van $3.6 - 1 = 2.6$ tot het expertgemiddelde en behaalt in dit geval een score van slechts $5 - 2.6 = 2.4$. Zodoende scoort de kandidaat hoger op de VR-SJT naarmate de antwoorden dichterbij dat van het expertgemiddelde liggen. De totaalscore op de VR-SJT werd berekend door de scores van alle items bij elkaar op te tellen.

Beschrijvende resultaten

Voorafgaand aan de hypothesetoetsing zijn verbanden tussen alle variabelen onderzocht. De assumptie van normaliteit werd geschonden voor de VR-SJT totaalscore en beide subschaalscores van de VR-SJT, en ook voor de video-SJT totaalscore en de subschaalscores van de video-SJT, waardoor er gebruik is gemaakt van een verdelingsvrije toets – in dit geval Spearman's rangcorrelatiecoëfficiënt. Naast de rangcorrelatiecoëfficiënten zullen ook de 95% betrouwbaarheidsintervallen (*confidence interval*, CI) worden vermeld. Tabel 1 geeft de correlaties weer tussen alle bestudeerde variabelen, evenals de gemiddelden en standaarddeviaties van de variabelen.

Tabel 1 laat zien dat enkele demografische kenmerken significant correleren met de bestudeerde variabelen. Genderverschillen werden gevonden voor de kandidaatperceptie van de VR-SJT, $r_s(94) = .27, p = .008, 95\% \text{ CI } [.08, .46]$ waarbij mannen de VR-SJT als positiever ervaren. Het opleidingsniveau correleerde positief met de totaalscore op de VR-SJT, $r_s(91) = .25, p = .015, 95\% \text{ CI } [.06, .44]$. Ook correleerde opleidingsniveau positief met de subschaal taakgericht leidinggeven van de VR-SJT, $r_s(91) = .24, p = .021, 95\% \text{ CI } [.03, .44]$ en met de subschaal persoonsgericht leidinggeven voor de VR-SJT, $r_s(91) = .27, p = .010, 95\% \text{ CI } [.07, .46]$. Deze resultaten betekenen dat hoger opgeleiden hoger scoorden op de VR-SJT. Het opleidingsniveau liet daarnaast een negatieve relatie zien met de kandidaatperceptie van de VR-SJT, $r_s(87) = -.25, p = .022, 95\% \text{ CI } [-.44, -.05]$. Met andere woorden; hoe hoger opgeleid hoe minder positief de VR-SJT beoordeeld werd.

De kandidaatperceptie van de VR-SJT correleerde tot slot opvallend genoeg negatief met de totaalscore op de VR-SJT, $r_s(94) = -.44, p < .001, 95\% \text{ CI } [-.57, -.28]$, met de subschaal taakgericht leidinggeven, $r_s(94) = -.39, p < .001, 95\% \text{ CI } [-.54, -.23]$ en met de subschaal persoonsgericht leidinggeven, $r_s(94) = -.43, p < .001, 95\% \text{ CI } [-.59, -.26]$.

Tabel 1.
Beschrijvende resultaten en Spearman's rangcorrelatiecoëfficiënten tussen alle variabelen

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
<i>Demografische variabelen</i>																							
1. Leeftijd	43.04	10.96	-																				
2. Geslacht	.48	.50	.07	-																			
3. Opleidingsniveau	1.38	.59	-.23	-.01	-																		
<i>VR-SJT^a</i>																							
4. VR-SJT totaal	328.61	14.24	-.05	-.16	.25*	(.87)																	
5. VR-SJT taakgericht	114.89	5.72	-.05	-.12	.24*	.87**	(.71)																
6. VR-SJT persoonsgericht	123.28	5.99	-.01	-.12	.27**	.83**	.62**	(.73)															
<i>Video-SJT^b</i>																							
7. SQ-leid totaal	294.58	8.36	.05	-.15	.10	.43**	.32*	.33*	(.82)														
8. SQ-leid taakgericht	137.57	4.77	.15	-.04	.09	.33*	.36**	.19	.85**	(.72)													
9. SQ-leid persoonsgericht	157.01	4.60	-.07	-.17	.15	.39**	.22	.34*	.84**	.47**	(.70)												
<i>Cognitieve capaciteiten^c</i>																							
10. ICCT-HV totaal	55.29	11.35	-.12	.16	.32*	.16	.10	.14	.27	.27	.31*	(.85)											
11. ICCT-HV verbaal	20.60	5.14	.13	.14	.33*	.23	.20	.18	.34*	.35*	.34*	.81**	(.74)										
12. ICCT-HV numeriek	17.00	4.89	-.19	.22	-.03	.04	.03	.02	-.03	.06	.01	.75**	.38**	(.80)									
13. ICCT-HV ruimtelijk	17.69	4.13	-.37*	.04	.41**	.15	.04	.15	.21	.13	.31	.82**	.51**	.56**	(.75)								
<i>Persoonlijkheid^d</i>																							
14. Integriteit	129.30	11.89	.13	-.07	.21	-.20	-.27	-.18	-.03	-.07	.05	.12	.10	.09	.11	(.91)							
15. Emotionaliteit	91.10	12.00	.18	-.22	.07	.10	-.04	.13	.11	.01	.16	-.04	-.04	-.10	-.10	.30*	(.88)						
16. Extraversie	125.92	10.34	.07	-.07	-.07	-.39**	-.22	-.35*	-.30*	-.10	-.40**	-.26	-.25	-.08	-.31*	-.06	.23	(.90)					
17. Verdraagzaamheid	102.14	9.75	.28	.02	.29*	-.18	-.12	-.08	-.15	-.08	-.16	-.01	-.02	.05	-.06	.50**	-.02	.07	(.88)				
18. Consciëntieusheid	119.30	11.51	.08	-.04	-.08	-.03	-.15	-.14	.15	.11	.11	-.15	-.09	-.17	-.18	.03	.12	.26	-.14	(.85)			
19. Openheid voor ervaringen	113.20	13.73	.19	-.19	.24	.13	.10	.06	.18	.13	.11	-.05	.14	-.23	-.08	-.03	-.24	.21	-.06	.34*	(.87)		
<i>Kandidaatperceptie</i>																							
20. VR-SJT kandidaatperceptie ^e	4.36	.39	-.16	.27**	-.25*	-.44**	-.39**	-.43**	-.03	-.03	.01	.05	.05	.07	.02	.05	-.14	.28	-.06	.08	-.10	(.66)	
21. SQ-leid kandidaatperceptie ^f	3.97	.43	.19	.21	-.14	-.03	.06	-.04	-.01	-.02	.07	.30	.23	.35*	.16	.32*	.23	.08	.19	-.14	-.17	.27	(.70)

Noot. Betrouwbaarheidscoëfficiënten zijn weergegeven in de diagonaal. Geslacht is als volgt gecodeerd: 0 = vrouw, 1 = man. Opleidingsniveau is als volgt gecodeerd: 0 = (v)mbo, havo en vwo, 1 = hbo, 2 = wo. De schaalscores zijn als volgt: VR-SJT totaal (min = 191, max = 378), VR-SJT taakgericht (min = 67, max = 132), VR-SJT persoonsgericht (min = 72, max = 142), SQ-leid totaal (min = 139, max = 310), SQ-leid taakgericht (min = 67, max = 150), SQ-leid persoonsgericht (min = 41, max = 170), ICCT-HV totaal (min = 0, max = 100), ICCT-HV verbaal, numeriek en ruimtelijk (min = 0, max = 35), HEXACO persoonlijkheidskenmerken (min = 32, max = 160), VR en SQ kandidaatperceptie (min = 1, max = 5).

^aN = 101. ^bN = 57. ^cN = 45. ^dN = 50. ^eN = 94. ^fN = 54.

* $p < .05$. ** $p < .01$. (tweezijdig getoetst)

Hypothesetoetsing

De eerste hypothese stelde dat de VR-SJT een beroep zou doen op de constructen taakgericht leidinggeven en persoonsgericht leidinggeven. Drie ervaren experts (twee arbeidpsychologen en een assessmentspecialist) hebben onafhankelijk van elkaar de items onderverdeeld in ofwel taak- of persoonsgericht leidinggeven. Vervolgens werd een *one-way random effects* intra-klasse correlatiecoëfficiënt (ICC) voor absolute overeenstemming berekend. De beoordelaarsovereenstemming was hoog (ICC = .88), waardoor de inhoudsvaliditeit lijkt aangetoond. Daarnaast is de convergente validiteit onderzocht. De subschaal VR-SJT taakgericht leidinggeven correleerde significant met de subschaal video-SJT taakgericht leidinggeven, $r_s(57) = .36, p = .006, 95\% \text{ CI } [.10, .58]$ en de subschaal VR-SJT persoonsgericht leidinggeven correleerde met de subschaal video-SJT persoonsgericht leidinggeven, $r_s(57) = .34, p = .010, 95\% \text{ CI } [.08, .55]$. Verder is ook de discriminante validiteit onderzocht. De subschaal VR-SJT taakgericht correleerde niet significant met de subschaal video-SJT persoonsgericht zoals te zien in Tabel 1, $r_s(57) = .22, p = .108, 95\% \text{ CI } [-.05, .45]$. Vice versa gold hetzelfde, de subschaal VR-SJT persoonsgericht correleerde niet significant met de subschaal video-SJT taakgericht $r_s(57) = .19, p = .149, 95\% \text{ CI } [-.08, .43]$. De eerste hypothese kan voorzichtig worden ondersteund.

Consistent met hypothese 2, die stelde dat de VR-SJT een positieve gemiddelde correlatie zou vertonen met de video-SJT, werd er een significante correlatie gevonden tussen de totaalscore VR-SJT en de totaalscore video-SJT van gemiddeld tot sterke grootte, $r_s(57) = .43, p = .001, 95\% \text{ CI } [.11, .62]$.

De derde hypothese stelde dat de constructen taakgericht leidinggeven en persoonsgericht leidinggeven in de VR-SJT positief gemiddeld met elkaar zouden correleren. Er werd een correlatie verwacht tussen de .30 en .50. De correlatie tussen de constructen bleek sterker dan verwacht, waardoor deze hypothese niet kan worden ondersteund, $r_s(101) = .62, p < .001, 95\% \text{ CI } [.45, .74]$.

In lijn met hypothesen 4a, 4b en 4c, die stelden dat de VR-SJT over een betrouwbaarheid van ten minste .70 zou moeten beschikken voor de totaalscore alsmede voor de constructen taakgericht leidinggeven en persoonsgericht leidinggeven, werden voldoende hoge betrouwbaarheidscoëfficiënten gevonden. Tabel 1 laat zien dat $\alpha = .87$ voor de totaalscore van de VR-SJT. Voor het construct taakgericht leidinggeven werd gevonden dat $\alpha = .71$ en voor het

construct persoonsgericht leidinggeven werd gevonden dat $\alpha = .73$. De betrouwbaarheid van de verschillscore tussen VR taakgericht en VR persoonsgericht was echter laag, $r_{DD'} = .26$ (zie Drenth & Sijtsma, 2006, p. 241 voor de formule). Dit werd veroorzaakt door de sterke correlatie ($r = .62$) tussen beide subschalen.

Hypothese vijf stelde dat de VR-SJT een hogere correlatiecoëfficiënt met cognitieve capaciteiten zou vertonen dan met persoonlijkheidskenmerken. In tegenstelling tot de verwachtingen, correleerde de VR-SJT niet significant met cognitieve capaciteiten, $r_s(45) = .16$, $p = .285$, 95% CI [-.16, .46]. Van de zes persoonlijkheidskenmerken correleerde de VR-SJT alleen significant (negatief) met extraversie, $r_s(50) = -.39$, $p = .005$, 95% CI [-.61, -.13]. Er werden vervolgens zes Fisher z-transformaties uitgevoerd om te onderzoeken of de correlaties tussen de VR-SJT en cognitieve capaciteiten en de VR-SJT en persoonlijkheid van elkaar verschilden. Dit bleek bij slechts twee correlaties het geval. Ten eerste verschilde de correlatie tussen de VR-SJT en cognitieve capaciteiten ($r_s(45) = .16$, $p = .285$, 95% CI [-.16, .46]) en tussen de VR-SJT en integriteit ($r_s(50) = -.20$, $p = .167$, 95% CI [-.47, .10]) significant van elkaar ($z = 1.78$, $p = .038$). Ten tweede werd gevonden dat de correlatie tussen VR-SJT en cognitieve capaciteiten ($r_s(45) = .16$, $p = .285$, 95% CI [-.16, .46]) en tussen VR-SJT en extraversie ($r_s(50) = -.39$, $p = .005$, 95% CI [-.61, -.13]) significant van elkaar verschilde ($z = 2.33$, $p = .01$). Hoewel de correlatie tussen de VR-SJT en cognitieve capaciteiten niet significant was, blijkt uit deze resultaten dat de VR-SJT een hogere correlatiecoëfficiënt vertoont met cognitieve capaciteiten dan met de persoonlijkheidskenmerken integriteit en extraversie. De vijfde hypothese kan daarom deels worden ondersteund.

De zesde hypothese stelde dat de VR-SJT positiever beoordeeld zou worden dan de video-SJT. Een gepaarde t-toets toonde dat de participanten de VR-SJT positiever beoordeelden ($M = 4.40$, $SD = 0.38$) dan de video-SJT ($M = 3.96$, $SD = 0.43$, $t(51) = 6.31$, $p < .001$). Het gevonden effect was groot, $d = 0.87$. Op basis van dit resultaat kan de zesde hypothese worden ondersteund. Om te kijken waarom de VR-SJT positiever werd beoordeeld, zijn de items uit de kandidaatperceptie vragenlijst apart bestudeerd. Tabel 2 toont de resultaten van de gepaarde t-toetsen voor ieder item van de kandidaatperceptie vragenlijst. Kandidaten vonden de VR-SJT leuker om te doen, $t(56) = 3.46$, $p = .001$, $d = 0.67$, vonden de VR-SJT nieuwer en moderner, $t(55) = 11$, $p < .001$, $d = 1.67$ en deden beter hun best om de VR-SJT te maken, $t(56) = 2.70$, $p = .009$, $d = 0.40$.

Tabel 2.

Resultaten van gepaarde t-toets en beschrijvende resultaten voor kandidaatperceptie van de VR-SJT en de video-SJT

Vraag	VR-SJT		video-SJT		t-toets	d	df
	M	SD	M	SD			
Ik vond het leuk om deze test te doen	4.53	.60	4.04	.84	3.46*	0.67	56
Het was me duidelijk wat er in deze test van mij werd verwacht	4.47	.63	4.45	.63	ns	0.03	54
De inhoud van de test was realistisch en herkenbaar	4.21	.84	4.21	.67	ns	0	56
Deze test komt op mij over als nieuw en modern	4.48	.71	3.09	.98	11**	1.62	55
Deze test lijkt mij relevant voor het voorspellen van leidinggevende vaardigheden	4.00	.61	3.95	.78	ns	0.07	54
Ik heb mijn best gedaan om deze test goed te maken	4.74	.44	4.53	.60	2.70*	0.40	56

Noot. Minimumscore per vraag = 1, maximumscore per vraag = 5.

* $p < .01$, ** $p < .001$

Discussie

In dit onderzoek stond de ontwikkeling van een VR-SJT centraal. Voor zover bekend is dit de eerste test die VR technologie gebruikt om een SJT te construeren. Het instrument is ontworpen om het inzicht van een kandidaat in effectief leiderschap te meten. Uniek aan VR-technologie is dat situaties meer realistisch en waarheidsgetrouw kunnen worden weergegeven. Daarnaast worden deelnemers volledig ondergedompeld in de virtuele wereld, waardoor zij niet afgeleid kunnen raken door externe prikkels. Experts op het gebied van leiderschap werden getest om de normgroep te creëren, waarna assessmentkandidaten en overige deelnemers de VR-SJT hebben gemaakt om het instrument verder te valideren. Centraal stond de vraag of er voldoende psychometrische basis zou worden gevonden voor een SJT in VR format. Er werd verwacht dat de VR-SJT een beroep zou doen op de concepten taakgericht leidinggeven en persoonsgericht leidinggeven, en dat de VR-SJT op betekenisvolle wijze zou samenhangen met een video-SJT met een vergelijkbare meetpretentie. Daarnaast werd verwacht dat het instrument

een sterkere relatie zou vertonen met cognitieve capaciteiten dan met persoonlijkheid. Tot slot is de kandidaatperceptie van beide SJTs onderzocht, waar er verwacht werd dat de VR-SJT positiever beoordeeld zou worden dan de video-SJT.

Judge et al. (2004) tonen in hun meta-analyse aan hoe (effectieve) leiderschapuitkomsten zijn terug te leiden naar het soort gedrag van de leidinggevende. Uit dit onderzoek bleek dat voornamelijk taakgericht en persoonsgericht leidinggevend gedrag in belangrijke mate de werkprestatie, motivatie en tevredenheid van de ondergeschikten en van het team voorspelt. Het is daarom belangrijk dat toekomstige leidinggevers voldoende kennis en inzicht hebben in deze gedragingen en dat zij (in)effectieve voorbeelden van zowel taak- als persoonsgericht leidinggeven correct kunnen herkennen en kunnen beoordelen. De VR-SJT beoogt daarom een beroep te doen op deze twee constructen zodat er getoetst kan worden of toekomstige leidinggevers over voldoende leidinggevend inzicht beschikken. In lijn met de eerste hypothese, bieden de resultaten van het huidige onderzoek voorzichtig steun voor het bestaan van deze twee constructen binnen de VR-SJT. Dit blijkt uit de bevinding dat convergente en discriminante validiteit konden worden aangetoond voor de constructen taakgericht en persoonsgericht leidinggeven tussen de VR-SJT en de video-SJT, die een vergelijkbare meetpretentie heeft. Daarnaast blijkt er een hoge mate van overeenstemming tussen een drietal experts die de items van de VR-SJT onafhankelijk van elkaar hebben ingedeeld in de categorieën taakgericht en persoonsgericht leidinggeven. Er moet echter wel rekening worden gehouden met de relatief sterke correlatie tussen de VR-SJT schalen taak- en persoonsgericht leidinggeven onderling ($r = .62$). Zo'n sterk verband kan namelijk doen vermoeden dat de schalen hetzelfde construct meten. Tot op zekere hoogte doen ze dat ook; effectief leiderschap is immers vaak situationeel van aard en een effectieve leider zal beide soorten gedragingen moeten kunnen beheersen in dezelfde context (Derue et al., 2011). Het feit dat de VR-SJT zich afspeelt in een vergaderomgeving moet ook in ogenschouw worden genomen. Odermatt et al. (2017) halen bijvoorbeeld het punt aan dat taak- en persoonsgericht vergadergedrag vaak samenhangen. Zij noemen het voorbeeld van een leidinggevende die een groepsconflict oplost door iedere deelnemer van de vergadering een beurt te geven om te spreken. In dit voorbeeld structureert de leidinggevende de vergadering (taakgericht gedrag) met als doel om ieder lid eerlijk te behandelen en even veel spreektijd te geven (persoonsgericht gedrag). De sterke correlatie tussen de constructen kan ook een gevolg zijn van de aard van de test. SJTs en constructvaliditeit

hebben namelijk een moeizame verhouding met elkaar (McDaniel et al., 2016). Een minpunt van SJTs is bijvoorbeeld dat ze veelal heterogeen zijn op itemniveau, zodat ieder item een beroep doet op meerdere, onderling niet gerelateerde, constructen. Factoranalyse van SJTs is dan ook vaak problematisch en levert tal van factoren op die lastig zijn te interpreteren (Lievens, 2017; Lievens et al., 2008). De constructvaliditeit van SJTs is statistisch gezien dus lastig te bepalen. De hoge mate van overeenstemming tussen de experts bij het indelen van de items levert in het huidige onderzoek dan ook een belangrijke bijdrage aan de steun voor de eerste hypothese in de vorm van inhoudsvaliditeit.

Ondersteuning wordt ook gevonden voor de tweede hypothese, die stelde dat de VR-SJT en de video-SJT positief zouden correleren, met een correlatie van gemiddelde grootte. Dit gevonden resultaat demonstreert de zogeheten soortgenootvaliditeit van het instrument. De VR-SJT laat zien op betekenisvolle wijze samen te hangen met de video-SJT, terwijl er voldoende unieke variantie overblijft. Deze bevinding biedt daarnaast de gelegenheid voor vervolgonderzoek om de mogelijke incrementele validiteit van de VR-SJT boven de video-SJT na te gaan.

Zoals eerder aangegeven, correleerden de VR-SJT subschalen taak- en persoonsgericht leidinggeven relatief sterk met elkaar. Dit resultaat is niet in lijn met de verwachting van hypothese drie, die een positieve correlatie van gemiddelde grootte veronderstelde. Deze relatie kan de interpretatie van de constructvaliditeit van het instrument bemoeilijken, zoals hierboven werd beschreven.

In tegenstelling tot vele andere SJTs, werden er in het huidige onderzoek voldoende hoge betrouwbaarheidscoëfficiënten gevonden voor de VR-SJT totaalscore en voor beide subschaalscores, zoals in hypothese vier werd verwacht. Catano et al. (2012), bijvoorbeeld, vinden in hun meta-analyse van 39 SJT studies een betrouwbaarheidscoëfficiënt van slechts $\alpha = .46$. De lage betrouwbaarheid van SJTs wordt in de literatuur dan ook als een van de grootste kritiekpunten genoemd. De betrouwbaarheid van $\alpha = .87$ die voor de VR-SJT totaalscore gevonden werd, staat dan ook in sterk contrast met dergelijke bevindingen. Dit resultaat kan mogelijk verklaard worden door de relatieve homogeniteit van de meetpretentie ‘inzicht in effectief leiderschap’, ten opzichte van andere meer heterogene concepten zoals interpersoonlijke vaardigheden en persoonlijkheid die ook vaak met behulp van SJTs worden gemeten (Christian et al., 2010). Daarnaast is in de ontwikkelfase van de VR-SJT voor een meer construct-gedreven

methode gekozen om het instrument te ontwikkelen. Voor het creëren van de items zijn bijvoorbeeld kritieke incidenten verzameld binnen het raamwerk van taak- en persoonsgericht leidinggeven. Op die manier zijn er effectieve en minder effectieve incidenten verzameld van zowel taak- als persoonsgericht leidinggevend gedrag, waar de traditionele manier van kritieke incidenten verzamelen zich minder richt op de achterliggende constructen. In voorkomende gevallen worden effectieve en minder effectieve incidenten verzameld van leidinggevend gedrag in het algemeen, zonder nadrukkelijk rekening te houden met een theoretisch raamwerk of achterliggend construct. Meerdere onderzoekers pleiten de laatste jaren voor een meer construct-gedreven opbouw van SJTs, omdat dit de constructvaliditeit en de betrouwbaarheid van SJTs ten goede komt (Guenole et al., 2017; Lievens, 2017). Daarmee voldoet de VR-SJT voor een deel aan de aanbevelingen die door deze onderzoekers wordt gedaan, wat een positieve bijdrage levert aan het onderzoeksveld van de SJT.

De vijfde hypothese stelde dat de VR-SJT een hogere correlatiecoëfficiënt met cognitieve capaciteiten zou vertonen dan met persoonlijkheidskenmerken. Hoewel er in het huidige onderzoek geen significant verband werd gevonden tussen de scores op de VR-SJT en cognitieve capaciteiten, liet een Fisher z-transformatie zien dat de correlatie tussen VR-SJT en cognitieve capaciteiten sterker was dan de correlatie tussen VR-SJT en de persoonlijkheidskenmerken integriteit en extraversie. De vijfde hypothese kan daarom slechts deels worden ondersteund, al dient dit resultaat voorzichtig te worden geïnterpreteerd omdat de correlatie tussen de scores op de VR-SJT en cognitieve capaciteiten niet significant bleek. Dit is opmerkelijk, omdat de VR-SJT inzicht in effectief leiderschap meet – en dus een bepaalde vorm van kennis toetst. Daarnaast bestaat het instrument voor een groot deel uit op kennisgeoriënteerde items, een type responsinstructie die doorgaans een correlatie met cognitieve capaciteiten laat zien (McDaniel et al., 2007; Ostrom et al., 2015). Het is daarbij wel belangrijk om zich te realiseren dat slechts 45 deelnemers de cognitieve capaciteitentest hebben ondergaan. Het is daarom aannemelijk dat er onvoldoende onderscheidingsvermogen (*power*) is geweest om een significante relatie te ontdekken. Opvallend is verder het gemiddelde tot sterke negatieve verband tussen de VR-SJT totaalscore en het persoonlijkheidskenmerk extraversie. Ook correleerde extraversie negatief met de subschaal persoonsgericht leidinggeven maar liet het geen verband zien met de subschaal taakgericht leidinggeven. Ditzelfde patroon werd overigens ook gevonden voor de correlatie tussen extraversie en de video-SJT. Wellicht overschatten extraverte mensen hun

beoordelingsvermogen in sociale situaties. De dimensie extraversie bestaat binnen de HEXACO persoonlijkheidsvragenlijst uit onder andere de facetten sociale zelfwaardering en sociale bravoure. Het kan dus zo zijn dat deelnemers die hoog scoren op deze facetten (te) veel zelfvertrouwen hebben in het beoordelen van de situaties en in hun beoordeling van de effectiviteit van leiderschap. Dit lijkt op het *too much of a good thing* principe, de theorie die een curvilineaire relatie veronderstelt tussen persoonlijkheid en werkprestatie (Le et al., 2011). Volgens dit principe is er een optimale waarde op persoonlijkheidsdimensies en neemt de effectiviteit van een persoonlijkheidsdimensie af naargelang deze te weinig of juist te sterk aanwezig is. Het persoonlijkheidskenmerk extraversie zou in dit geval te sterk aanwezig zijn, wat een nadelig effect kan hebben op het beoordelingsvermogen. Het zou ook kunnen zijn dat extraverte mensen de rol van observator in de VR-SJT moeilijker kunnen omarmen. Met een extraverte persoonlijkheid zou men wellicht actiever willen bijdragen aan de groepsvergadering die in de VR omgeving plaats vindt, waardoor zij zich minder op hun plaats voelen bij het observeren van het gedrag van de groepsleden en de leidinggevende. Dit kan de testmotivatie van deze groep deelnemers verlagen, wat op zijn beurt weer voor een lagere score op de VR-SJT kan zorgen. Het vormt een suggestie voor vervolgonderzoek om de precieze relatie tussen extraversie en inzicht in effectief leiderschap verder te onderzoeken.

Tot slot bevestigen de resultaten de zesde hypothese, die stelde dat deelnemers de VR-SJT positiever zouden beoordelen dan de video-SJT. De effectgrootte ($d = 0.87$) laat daarnaast zien dat het gevonden verschil groot is. Op itemniveau werd gevonden dat kandidaten de VR-SJT leuker vinden, dat ze de test nieuwer en moderner vinden en dat ze beter hun best hebben gedaan om de VR-SJT goed te maken. Daarbij moet wel vermeld worden dat de spreiding in de kandidaatperceptie vragenlijst klein is. Zowel de video-SJT als de VR-SJT worden namelijk zeer positief ervaren zoals te zien in Tabellen 1 en 2. Daarnaast moet er rekening worden gehouden met het feit dat de gepaarde t-toets op een kleine steekproef ($n = 52$) is uitgevoerd. Tot slot zal een kritische lezer opmerken dat het gemiddelde van de VR-SJT kandidaatperceptie vragenlijst omhoog getild zal zijn door het item “Deze test komt op mij over als nieuw en modern”. Een test die gebruikt maakt van VR technologie is per definitie nieuwer en moderner dan een test met alleen videobeelden. Echter, wanneer dit item uit de kandidaatperceptie vragenlijst wordt verwijderd, wordt de VR-SJT nog steeds als positiever beoordeeld dan de video-SJT, $t(52) = 2.61$, $p = .012$. De effectgrootte neemt daarmee wel af, van $d = 0.87$ tot $d = 0.42$. Ondanks de

bovengenoemde limitaties is dit een belangrijke bevinding. De perceptie van deelnemers speelt een belangrijke rol bij de testmotivatie en bij de acceptatie van het selectieproces (McCarthy et al., 2017). Indruksvalide en door kandidaten geaccepteerde instrumenten zorgen voor meer draagvlak onder de deelnemers, wat de testmotivatie verhoogt en op zijn beurt voor een hogere testprestatie kan zorgen (Anderson et al., 2010; McCarthy et al., 2017). In de meta-analyse van Anderson et al. (2010) wordt zelfs gevonden dat de kandidaatperceptie en de indruksvaliditeit van een instrument sterk correleren ($r = .66$ en $r = .63$) met de operationele validiteit van dat instrument, wat laat zien dat deelnemers een nauwkeurig oordeel kunnen vellen over de validiteit van een instrument.

Ondanks het feit dat er in het huidige onderzoek werd gevonden dat de testmotivatie hoger is voor de VR-SJT dan voor de video-SJT, werd er een negatieve correlatie gevonden tussen de kandidaatperceptie van de VR-SJT en de testprestatie. Deze gemiddeld tot sterke negatieve correlatie laat zien dat deelnemers die de VR-SJT positiever beoordelen, lager scoren op de VR-SJT en vice versa. Dit effect werd gevonden voor zowel de totaalscore als voor de schaalcores taak- en persoonsgericht leidinggeven. Er kan geen eenduidige verklaring worden gegeven voor deze bevindingen. Het zou zo kunnen zijn dat kandidaten die hoog scoren op de VR-SJT de test te gemakkelijk vinden. Hierdoor zouden zij het gevoel kunnen krijgen dat ze niet al hun kwaliteiten en/of leidinggevende vaardigheden kunnen demonstreren, waardoor zij de test minder positief beoordelen. Een andere verklaring ligt in het innoverende karakter van de test. Het is mogelijk dat kandidaten die de VR-SJT positief beoordelen zich in hun enthousiasme laten afleiden door de VR technologie. Het staat de kandidaat immers vrij om 360 graden om zich heen te kijken en om de aandacht te richten op ieder lid van de groepsvergadering. Het is dan dus mogelijk dat een kandidaat minder aandacht schenkt aan het gedrag van de leidinggevende, waardoor er minder items goed worden beantwoord en er een lagere testscore wordt behaald. In toekomstig onderzoek is het belangrijk om deze zaken in overweging te nemen. Het kan bijvoorbeeld helpen om kandidaten die nog geen eerdere ervaring met VR hebben opgedaan een korte proefafname te laten maken. Op die manier kunnen zij wennen aan de technologie en kunnen zij hun enthousiasme kwijt in een laagdrempelige omgeving, zonder dat dit ten koste gaat van hun testprestatie.

In het huidige onderzoek werd verder gevonden dat mannen de VR-SJT positiever beoordeelden dan vrouwen. Het kan zo zijn dat mannen meer affiniteit hebben met VR

technologie dan vrouwen. Geslacht toonde daarentegen geen relatie met de daadwerkelijke testcores, waardoor er geen aanwijzingen worden gevonden dat de VR-SJT niet eerlijk zou zijn voor verschillende geslachten. Het vormt voor toekomstig onderzoek wellicht een interessant punt om te onderzoeken hoe mannen en vrouwen verschillen in de mate van acceptatie van op VR gebaseerde tests.

Het laatste verschil in kandidaatperceptie van de VR-SJT hing af van iemands opleidingsniveau. Hoe hoger de kandidaat was opgeleid, hoe minder positief de VR-SJT werd beoordeeld. Wellicht kijken hoger opgeleiden kritischer naar de inhoud en naar de meetpretentie van het instrument. Het is wel belangrijk rekening te houden met het feit dat lager opgeleiden ondervertegenwoordigd zijn in dit onderzoek. Het verschil zit dan ook voornamelijk tussen hbo- en wo-opgeleide deelnemers.

Daarnaast bleek dat hoe hoger het opleidingsniveau, hoe hoger men scoorde op de VR-SJT. Een verklaring voor dit resultaat kan worden gevonden in het feit dat hoger opgeleiden wellicht vaker functies bekleden waarin zij (nauw) samenwerken met leidinggevenden of waar er meer wordt vergaderd. In voorkomende gevallen zouden zij meer zijn blootgesteld aan situaties waar (in)effectief leidinggevend gedrag wordt vertoond en zouden zij dus al een mentale representatie hebben van (in)effectief leiderschap. Het kan ook zo zijn dat dit resultaat een indirect effect is van iemands cognitieve capaciteiten. De score op de cognitieve capaciteitentest en het opleidingsniveau waren namelijk wel positief aan elkaar gerelateerd. Zoals eerder vermeld, werd er in het huidige onderzoek geen relatie gevonden tussen cognitieve capaciteiten en de scores op de VR-SJT. Slechts 45 deelnemers hebben echter de cognitieve capaciteitentest gemaakt, waardoor er wellicht onvoldoende *power* was om een significante relatie te ontdekken. Toekomstig onderzoek kan zich richten op de precieze relatie tussen opleidingsniveau en de testprestatie op de VR-SJT. Het kan dan verstandig zijn om de werkervaring en de hoeveelheid blootstelling aan leidinggeven en vergaderen mee te nemen in de analyses.

Beperkingen van het onderzoek en suggesties voor toekomstig onderzoek

Allereerst is het belangrijk te beseffen dat er aan de hand van de uitkomsten van het huidige onderzoek nog geen uitspraken kunnen worden gedaan over de predictieve validiteit van de VR-SJT. Hoewel de psychometrische eigenschappen van het instrument er veelbelovend uitzien, vormt het nog maar de vraag in hoeverre de score op de VR-SJT voorspellend is voor daadwerkelijk leidinggevend gedrag en/of werkprestatie in een leidinggevende functie. Dit vormt

dan ook de meest dringende suggestie om in toekomstig onderzoek nader te bestuderen. In het verlengde hiervan kan dan ook de incrementele validiteit worden bestudeerd ten opzichte van andere methoden. Gezien de beperkte overlap met cognitieve capaciteiten en persoonlijkheid lijkt hier voldoende ruimte voor te zijn, al zal toekomstig onderzoek dit empirisch moeten vaststellen.

Verder moet er rekening worden gehouden met de beperkte steekproefgrootte van dit onderzoek. Hoewel 101 deelnemers de VR-SJT hebben doorlopen, hebben slechts 57 deelnemers zowel de VR-SJT als de video-SJT kunnen maken. De cognitieve capaciteitentest en de persoonlijkheidsvragenlijst zijn, in combinatie met de VR-SJT, bij respectievelijk slechts 45 en 50 deelnemers afgenomen. De resultaten en conclusies met betrekking tot deze instrumenten moeten daarom met voorzichtigheid worden geïnterpreteerd. Hoewel de VR-SJT kandidaatperceptie vragenlijst door aanzienlijk meer kandidaten is ingevuld ($n = 94$), moeten ook deze resultaten met een slag om de arm worden geïnterpreteerd. De interne betrouwbaarheid ($\alpha = .66$) is niet bijzonder hoog en de vragenlijst meet een breed aantal concepten. Er kunnen daarom slechts in beperkte mate uitspraken worden gedaan over de daadwerkelijke indrukvaliditeit en de testmotivatie van de VR-SJT. Voor vervolgonderzoek is het aan te raden om afzonderlijke vragenlijsten te gebruiken die concepten zoals indrukvaliditeit en testmotivatie meten.

Noemenswaardig is verder het relatief sterke verband tussen de concepten taakgericht en persoonsgericht leidinggeven in de VR-SJT. Men kan zich afvragen in hoeverre deze soorten leidinggeven hetzelfde meten. Daarentegen wordt er een hoge mate van overeenstemming gevonden tussen experts die de items onafhankelijk van elkaar hebben ingedeeld in de categorieën taak- en persoonsgericht leidinggeven, wat een aanwijzing vormt voor de inhoudsvaliditeit van beide onderdelen van leidinggeven. Hoewel sterke verbanden tussen onderdelen van SJTs vaker worden gevonden, is het iets om rekening mee te houden bij de onderscheidbaarheid van de testcores.

Tevens kan een kritische lezer opmerken dat de VR-SJT niet alleen iemands inzicht in leiderschap meet, maar ook iemands voorkeur voor een bepaalde stijl van leidinggeven. Een deelnemer die zelf graag een meer persoonsgerichte stijl van leidinggeven ontvangt, zal de effectiviteit van deze stijl waarschijnlijk hoger beoordelen. Voor toekomstig onderzoek is het dan ook interessant om de *implicit trait policies* (Motowidlo et al., 2006; Oostrom et al., 2012) van een kandidaat nader te bestuderen.

Daarnaast vormt het een suggestie voor toekomstig onderzoek om het effect van VR-technologie in tests grondiger te onderzoeken. In het huidige onderzoek wordt bijvoorbeeld gevonden dat mannen de VR-SJT positiever beoordelen dan vrouwen. Het is informatief om dit verschil in meer detail te onderzoeken. Verder kan er ook gekeken worden naar het gevoel van *presence* en het effect daarvan op concentratie, testmotivatie en testprestatie. Het gevoel van *presence* is een psychologische staat waarbij VR-gebruikers het idee krijgen dat zij zich daadwerkelijk bevinden in, en interacteren met, een virtuele omgeving. De mate van *presence* is in het huidige onderzoek niet als zodanig bestudeerd en er kunnen daarom geen uitspraken worden gedaan over de ‘onderdompelende’ ervaring die juist zo uniek is voor VR.

Tot slot is er in het huidige onderzoek geen rekening gehouden met mogelijke etnische verschillen in VR-SJT scores. Er wordt aangeraden om dit in vervolgonderzoek wel te doen, om na te gaan of de VR-SJT even eerlijk is voor deelnemers ongeacht hun etniciteit.

Praktische implicaties

De resultaten van het huidige onderzoek en de psychometrische eigenschappen van de VR-SJT geven voorlopig voldoende vertrouwen voor een praktisch inzetbaar instrument om inzicht in effectief leiderschap te meten. Hoewel de voorspellende kracht nog moet worden aangetoond, is de VR-SJT breed toepasbaar, zoals de meeste SJTs. Het instrument kan bijvoorbeeld worden gebruikt bij (leiderschaps)ontwikkeling, trainingen, potentieelmetingen en loopbaankeuzes. De beperkte en/of afwezige verbanden, en dus geringe overlap, met cognitieve capaciteiten en persoonlijkheid impliceert dat de VR-SJT een goede aanvulling kan zijn op cognitieve capaciteitentests, persoonlijkheidsvragenlijsten en dergelijke instrumenten. De VR-SJT vormt een innovatieve aanwinst voor assessmentbureaus en andere HR-professionals en is tevens een manier om zich te onderscheiden van de concurrentie en om de selectieprocedure aantrekkelijker te maken voor kandidaten. De VR-SJT wordt gunstig beoordeeld door deelnemers en zal waarschijnlijk positief bijdragen aan de algehele ervaring van een assessment, training en/of selectieprocedure. Dit is belangrijk, omdat er aanwijzingen zijn dat kandidaten bij een aantrekkelijke procedure sneller zullen solliciteren bij een organisatie en de werkgever vaker zullen aanbevelen bij kennissen en relaties (Hausknecht et al., 2004; McCarthy et al., 2017).

Conclusie

In dit onderzoek stond de vraag centraal of er voldoende psychometrische basis zou worden gevonden voor een situationele beoordelingstest of zogeheten *situational judgment test*

(SJT) in *virtual reality* (VR) format. Voor zover bekend is dit het eerste onderzoek dat gebruikt maakt van VR-technologie om een SJT te ontwikkelen. De resultaten geven voldoende vertrouwen voor het bestaansrecht van de VR-SJT. De interne betrouwbaarheid van het instrument is hoog en de soortgenootvaliditeit met een reeds gevalideerde video-SJT kon worden aangetoond. Het instrument beoogt het inzicht in effectief leiderschap te meten en levert een totaalscore alsmede een score voor taak- en persoonsgericht leidinggeven. De onderscheidbaarheid van beide leiderschapsstijlen moet echter nog duidelijker worden onderbouwd. De VR-SJT wordt positief ontvangen door kandidaten, die de test bovendien positiever beoordelen dan een video-SJT. Na de succesvolle transitie van tekstuele SJTs naar video-SJTs, kan VR het volgende medium worden om de SJT aantrekkelijker, moderner en meer waarheidsgetrouw te maken.

Referenties

- Aguinis, H., Henle, C. A., & Beaty Jr, J. C. (2001). Virtual reality technology: A new tool for personnel selection. *International Journal of Selection and Assessment*, 9(1-2), 70-83. <https://doi.org/10.1111/1468-2389.00164>
- Anderson, N. (2003). Applicant and recruiter reactions to new technology in selection: A critical review and agenda for future research. *International Journal of Selection and Assessment*, 11(2-3), 121-136. <https://doi.org/10.1111/1468-2389.00235>
- Anderson, N., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, 18(3), 291-304. <https://doi.org/10.1111/j.1468-2389.2010.00512.x>
- Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. A., Schmertz, S. K., Zimand, E., & Calamaras, M. R. (2013). Virtual reality exposure therapy for social anxiety disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 81(5), 751-760. <https://doi.org/10.1037/a0033559>
- Backhaus, K., & Tikoo, S. (2004). Conceptualizing and researching employer branding. *Career Development International*, 9(5), 501-517. <https://doi.org/10.1108/13620430410550754>
- Bailenson, J. (2018). *Experience on demand: What virtual reality is, how it works, and what it can do*. Geraadpleegd op <https://books.google.nl/books?hl=nl&lr=&id=2fkqDwAAQBAJ>
- Campion, M. C., Ployhart, R. E., & MacKenzie Jr, W. I. (2014). The state of research on situational judgment tests: A content analysis and directions for future research. *Human Performance*, 27(4), 283-310. <https://doi.org/10.1080/08959285.2014.929693>
- Catano, V. M., Brochu, A., & Lamerson, C. D. (2012). Assessing the reliability of situational judgment tests used in high-stakes situations. *International Journal of Selection and Assessment*, 20(3), 333-346. <https://doi.org/10.1111/j.1468-2389.2012.00604.x>
- 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(1), 143-159. <http://dx.doi.org/10.1037/0021-9010.82.1.143>

- 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*(1), 83-117. <https://doi.org/10.1111/j.1744-6570.2009.01163.x>
- Crook, A. E., Beier, M. E., Cox, C. B., Kell, H. J., Hanks, A. R., & Motowidlo, S. J. (2011). Measuring relationships between personality, knowledge, and performance using single-response situational judgment tests. *International Journal of Selection and Assessment*, *19*(4), 363-373. <https://doi.org/10.1111/j.1468-2389.2011.00565.x>
- De Leng, W. E., Stegers-Jager, K. M., Husbands, A., Dowell, J. S., Born, M. P., & Themmen, A. P. N. (2017). Scoring method of a situational judgment test: Influence on internal consistency reliability, adverse impact and correlation with personality? *Advances in Health Sciences Education*, *22*(2), 243-265. <https://doi.org/10.1007/s10459-016-9720-7>
- De Meijer, L. A. L., Born, M. P., van Zielst, J., & van der Molen, H. T. (2010). Construct-driven development of a video-based situational judgment test for integrity. *European Psychologist*, *15*(3), 229-236. <http://dx.doi.org/10.1027/1016-9040/a000027>
- Derue, D. S., Nahrgang, J. D., Wellman, N. E. D., & Humphrey, S. E. (2011). Trait and behavioral theories of leadership: An integration and meta-analytic test of their relative validity. *Personnel Psychology*, *64*(1), 7-52. <https://doi.org/10.1111/j.1744-6570.2010.01201.x>
- De Vries, R. E., Ashton, M. C., & Lee, K. (2009). De zes belangrijkste persoonlijkheidsdimensies en de HEXACO persoonlijkheidsvragenlijst. *Gedrag en Organisatie*, *22*(3), 232-274. Verkregen van https://www.gedragenorganisatie.nl/scripts/shared/artikel_pdf.php?id=GO-22-3-4
- Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in Psychology*, *6*(26), 1-9. <https://doi.org/10.3389/fpsyg.2015.00026>
- Drenth, P. J. D., & Sijtsma, K. (2006). *Testtheorie: Inleiding in de theorie van de psychologische test en zijn toepassingen* (4e ed.). Houten, Nederland: Bohn Stafleu van Loghum.
- Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., & Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, *47*(14), 2393-2400. <http://doi.org/10.1017/S003329171700040X>

- Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, *18*(4), 694-734. <https://doi.org/10.5465/amr.1993.9402210155>
- Guenole, N., Chernyshenko, O. S., & Weekly, J. (2017). On designing construct driven situational judgment tests: Some preliminary recommendations. *International Journal of Testing*, *17*(3), 234-252. <https://doi.org/10.1080/15305058.2017.1297817>
- Hartmann, T., Wirth, W., Schramm, H., Klimmt, C., Vorderer, P., Gysbers, A., . . . Gouveia, F. (2015). The spatial presence experience scale (SPES). *Journal of Media Psychology*, *28*(1), 1-15. <https://doi.org/10.1027/1864-1105/a000137>
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*(3), 639-683. <https://doi.org/10.1111/j.1744-6570.2004.00003.x>
- Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology*, *89*(1), 36-51. <https://dx.doi.org/10.1037/0021-9010.89.1.36>
- Kandalajt, M. R., Didehbani, N., Krawczyk, D. C., Allen, T. T., & Chapman, S. B. (2013). Virtual reality social cognition training for young adults with high-functioning autism. *Journal of Autism and Developmental Disorders*, *43*(1), 34-44. <https://doi.org/10.1007/s10803-012-1544-6>
- Kanning, U. P., Grewe, K., Hollenberg, S., & Hadouch, M. (2006). From the subjects' point of view: Reactions to different types of situational judgment items. *European Journal of Psychological Assessment*, *22*(3), 168-176. <https://doi.org/10.1027/1015-5759.22.3.168>
- Le, H., Oh, I. S., Robbins, S. B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: Curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology*, *96*(1), 113. <https://doi.org/10.1037/a0021016>
- 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*(1), 51-69. <https://doi.org/10.1002/hrm.20145>
- Lievens, F. (2017). Construct-driven SJTs: Toward an agenda for future research. *International Journal of Testing*, *17*(3), 269-276. <https://doi.org/10.1080/15305058.2017.1309857>

- Lievens, F., & Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment*, *10*(4), 245-257. <https://doi.org/10.1111/1468-2389.00215>
- Lievens, F., & De Soete, B. (2011). Instrumenten om Personeel te Selecteren in de 21ste Eeuw: Onderzoek en Praktijk. *Gedrag en Organisatie*, *24*(1), 18-42. Verkregen van https://www.gedragenorganisatie.nl/scripts/shared/artikel_pdf.php?id=GO-24-1-18
- Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology*, *9*(1), 3-22. <https://doi.org/10.1017/iop.2015.71>
- Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, *37*(4), 426-441. <https://doi.org/10.1108/00483480810877598>
- 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*(5), 1181-1188. <https://doi.org/10.1037/0021-9010.91.5.1181>
- McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant perspectives during selection: A review addressing “So what?,” “What’s new?,” and “Where to next?”. *Journal of Management*, *43*(6), 1693-1725. <https://doi.org/10.1177/0149206316681846>
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb III, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, *60*(1), 63-91. <https://doi.org/10.1111/j.1744-6570.2007.00065.x>
- McDaniel, M. A., List, S. K., & Kepes, S. (2016). The “hot mess” of situational judgment test construct validity and other issues. *Industrial and Organizational Psychology*, *9*(1), 47-51. <https://doi.org/10.1017/iop.2015.115>
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Predicting job performance using situational judgment tests: A clarification of the literature. *Journal of Applied Psychology*, *86*(4), 730-740. <http://dx.doi.org/10.1037/0021-9010.86.4.730>
- Motowidlo, S. J., Crook, A. E., Kell, H. J., & Naemi, B. (2009). Measuring procedural knowledge more simply with a single-response situational judgment test. *Journal of Business and Psychology*, *24*(3), 281-288. <https://doi.org/10.1007/s10869-009-9106-4>

- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*(6), 640-647. <http://dx.doi.org/10.1037/0021-9010.75.6.640>
- 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*(4), 749-761. <http://dx.doi.org/10.1037/0021-9010.91.4.749>
- Odermatt, I., König, C. J., Kleinmann, M., Bachmann, M., Röder, H., & Schmitz, P. (2018). Incivility in meetings: Predictors and outcomes. *Journal of Business and Psychology, 33*(2), 263-282. <https://doi.org/10.1007/s10869-017-9490-0>
- Odermatt, I., König, C. J., Kleinmann, M., Nussbaumer, R., Rosenbaum, A., Olien, J. L., & Rogelberg, S. G. (2017). On leading meetings: Linking meeting outcomes to leadership styles. *Journal of Leadership & Organizational Studies, 24*(2), 189-200. <https://doi.org/10.1177/1548051816655992>
- Olson-Buchanan, J. B., & Drasgow, F. (2006). Multimedia situational judgment tests: The medium creates the message. In J. A. Weekley & R. E. Ployhart (Red.), *Situational judgment tests: Theory, measurement, and application* (pp. 253-278). Geraadpleegd op <https://books.google.nl/books?id=vz4zjg5jzTIC&pg>
- Oostrom, J. K., Born, M. P., Serlie, A. W., & van der Molen, H. T. (2011). A multimedia situational test with a constructed-response format. *Journal of Personnel Psychology, 10*(2), 78-88. <https://doi.org/10.1027/1866-5888/a000035>
- Oostrom, J. K., Born, M. P., Serlie, A. W., & van der Molen, H. T. (2012). Implicit trait policies in multimedia situational judgment tests for leadership skills: Can they predict leadership behavior? *Human Performance, 25*(4), 335-353. <https://doi.org/10.1080/08959285.2012.703732>
- Oostrom, J. K., De Soete, B., & Lievens, F. (2015). Situational judgment testing: A review and some new developments. In I. Nikolaou, & J. K. Oostrom (Red.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 172-189). Geraadpleegd op <https://books.google.nl/books?hl=nl&lr=&id=FGtKCAAQBAJ&oi>
- Oostrom, J. K., Van Der Linden, D., Born, M. P., & Van Der Molen, H. T. (2013). New technology in personnel selection: How recruiter characteristics affect the adoption of

- new selection technology. *Computers in Human Behavior*, 29(6), 2404-2415. <https://doi.org/10.1016/j.chb.2013.05.025>
- Op de Beek, M. R., Oostrom, J. K., & Born, M. P. (2011). De webcamtest als voorspeller van professioneel gedrag. *Gedrag en Organisatie*, 24(3), 257-285. Verkregen van https://www.gedragenorganisatie.nl/scripts/shared/artikel_pdf.php?id=GO-24-3-257
- Patterson, F., Zibarras, L., Carr, V., Irish, B., & Gregory, S. (2011). Evaluating candidate reactions to selection practices using organisational justice theory. *Medical Education*, 45(3), 289-297. <https://doi.org/10.1111/j.1365-2923.2010.03808.x>
- Richman-Hirsch, 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(6), 880-887. <https://doi.org/10.1037/0021-9010.85.6.880>
- Rothbaum, B. O., Price, M., Jovanovic, T., Norrholm, S. D., Gerardi, M., Dunlop, B., . . . Ressler, K. J. (2014). A randomized, double-blind evaluation of D-cycloserine or alprazolam combined with virtual reality exposure therapy for posttraumatic stress disorder in Iraq and Afghanistan War veterans. *American Journal of Psychiatry*, 171(6), 640-648. <https://doi.org/10.1176/appi.ajp.2014.13121625>
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274. <https://doi.org/10.1037/0033-2909.124.2.262>
- Sharma, S., Gangopadhyay, M., Austin, E., & Mandal, M. K. (2013). Development and validation of a situational judgment test of emotional intelligence. *International Journal of Selection and Assessment*, 21(1), 57-73. <https://doi.org/10.1111/ijjsa.12017>
- Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016). Validity and reliability of situational judgement test scores: A new approach based on cognitive diagnosis models. *Organizational Research Methods*, 19(3), 506-532. <https://doi.org/10.1177%2F1094428116630065>
- Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, 81(2), 134-141. <https://doi.org/10.1037/0021-9010.81.2.134>
- Stemler, S. E., & Sternberg, R. J. (2006). Using situational judgment tests to measure practical intelligence. In J. A. Weekley & R. E. Ployhart (Red.), *Situational judgment*

- tests: Theory, measurement, and application* (pp. 107-131). Geraadpleegd op <https://books.google.nl/books?id=vz4zjg5jzTIC>
- Van der Maesen | Koch HRM-advies. (2011a). *Handleiding SQ-leidinggeven*. Verkregen van https://www.t-station.nl/upload/files/handleiding_sqleidinggeven_september_2011.pdf
- Van der Maesen | Koch HRM-advies. (2011b). *Internet cognitieve capaciteitentest hoog niveau*. Verkregen van <https://www.t-station.nl/internet-cognitieve-capaciteiten-test-hoog-icct-hv-nl-e-116.html>
- Van Gelder, J. L., de Vries, R. E., Demetriou, A., van Sintemaartensdijk, I., & Donker, T. (2019). The virtual reality scenario method: Moving from imagination to immersion in criminal decision-making research. *Journal of Research in Crime and Delinquency*, 56(3), 451-480. <https://doi.org/10.1177%2F0022427818819696>
- Van Gelder, J. L., Martin, C., Van Prooijen, J. W., De Vries, R., Marsman, M., Averdijk, M., . . . Donker, T. (2018). Seeing is believing? Comparing negative affect, realism and presence in visual versus written guardianship scenarios. *Deviant Behavior*, 39(4), 461-474. <https://doi.org/10.1016/j.chb.2017.11.036>
- Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, 19(3), 188-202. <https://doi.org/10.1016/j.hrmr.2009.03.007>