

Het objectief voorspellen van werkprestaties:

De mogelijkheden van tekstanalyse.

Jasper Stooker

Studentnummer: 2150670

Begeleider: Drs. D. J. Holtrop

Tweede beoordelaar: dr. J. K. Oostrom

Masterthese Arbeids- en Organisationspsychologie

Datum: 23-10-2014

Vrije Universiteit Amsterdam

Voorwoord

Voor U ligt mijn masterthese welke onderdeel is van de master Arbeids- en Organisationspsychologie aan de Vrije Universiteit te Amsterdam. Graag wil via deze weg alle mensen bedanken die op enigerlei wijze hebben bijgedragen aan de totstandkoming van mijn onderzoek naar een nieuwe assessment technologie. Allereerst wil ik Marcel Bechger bedanken voor de mogelijkheid om dit onderzoek bij Fluor B.V. in Haarlem te verrichten. Mijn dank gaat ook uit naar Ward van Breda die speciaal voor dit onderzoek een nieuw algoritme, genaamd Sentimentics, beschikbaar heeft gesteld. Daarnaast wil ik Janneke Oostrom en Reinout de Vries bedanken voor het meedenken over ideeën voor dit onderzoek en het beschikbaar stellen van materiaal. Tevens wil ik mijn scriptiebegeleider Djurre Holtrop bedanken voor het begeleiden en het commentariëren van mijn masterthese. Tenslotte gaat mijn dank uit naar mijn ouders en mijn vriendin die altijd zorgden voor de juiste motivatie en mij geholpen hebben met het uitschrijven van de spraaktaken.

Amsterdam, oktober 2014

Jasper Stooker

Samenvatting

In deze masterthesis is onderzocht in hoeverre het geautomatiseerd beoordelen van (gesproken) tekst beoordelingen door middel van vragenlijsten kan vervangen of aanvullen als voorspeller van werkprestatie. Dit onderzoek is uitgevoerd bij een groot ingenieursbureau met Nederlandstalige werknemers in de leeftijd van 21 tot 65. Voor dit onderzoek zijn bij 103 werknemers de HEXACO-60 persoonlijkheidsvragenlijst en een speciaal ontwikkelde spraaktaak afgenomen. De spraaktaak had als doel om de persoonlijkheidsdimensie Consciëntieusheid te meten. Als eerste is de constructvaliditeit aangetoond door de samenhang tussen de HEXACO en de spraaktaak te meten. Zoals verwacht was er een significant verband tussen Consciëntieusheid volgens de HEXACO-PI en volgens de spraaktaken ($r = .33, p < .01$), hoewel deze relatie zwakker is dan van identieke constructen verwacht wordt. Daarna is de predictieve validiteit (voor werkprestatie) van de geautomatiseerde beoordeling getoetst en vergeleken met de predictieve validiteit van de HEXACO. Er is een verband gevonden tussen HEXACO Consciëntieusheid zelfbeoordeling en Algemene werkprestatie ($r = 0.42, p < .01$). Tot slot is er een relatie gevonden tussen persoonlijkheid, gemeten met de spraaktaak, en Algemene werkprestatie ($r = 0.26, p < .05$). Echter, deze relatie heeft geen incrementele waarde boven de HEXACO zelfbeoordeling. Kortom, het geautomatiseerd beoordelen van persoonlijkheid om werkprestatie te voorspellen is dus mogelijk, maar in het huidige onderzoek wordt tevens aangetoond dat geautomatiseerde beoordelingen nog niet zelfbeoordelingen kunnen vervangen.

Inhoudsopgave

INLEIDING	5
Wetenschappelijke en praktische relevantie	7
THEORETISCH KADER	9
Persoonlijkheid en werkprestatie.....	9
Persoonlijkheid meten	11
Nieuwe assessmenttechnologie	13
ONDERZOEKSMETHODE	20
Onderzoeksdesign	20
Procedure	20
Participanten.....	21
Meetinstrumenten.....	22
RESULTATEN	29
Hypothese 1	29
Hypothese 2	30
Hypothese 3	31
Post-hoc analyses	33
DISCUSSIE	37
Theoretische en praktische implicaties	40
Beperkingen en suggesties voor vervolgonderzoek	42
CONCLUSIE	46
REFERENTIES	47
Bijlage 1. Factoranalyses	54
Bijlage 2. HEXACO-PI-60.....	56
Bijlage 3. Spraaktaak.....	62
Bijlage 4. Beoordelingsformulier Algemene werkprestatie	64
Bijlage 5. Confliction Value.....	66

Inleiding

Mensen zijn niet volledig rationeel in het beoordelen van anderen (Kerr, MacCoun, & Kramer, 1996). Veel onderzoek op het gebied van de arbeids- en organisatiepsychologie is besteed aan het aantonen van wat waarschijnlijk een niet-verrassende stelling is: menselijke oordelen zijn niet perfect. Wat deze onderzoeken naar menselijke oordelen nuttig en interessant maakt, is dat deze onvolkomenheden vaak meer ten grondslag liggen aan psychologische processen, dan aan willekeurige schommelingen rond rationele, voorgeschreven of ideale oordelen. Sterker nog, mensen vertonen continu *cognitive biases* in hun oordelen. Een *cognitive bias* is een patroon van afwijking in het oordeel, waarbij conclusies worden getrokken over mensen en situaties gebaseerd op onlogische gedachtes (Haselton, Nettle, & Andrews, 2005). Deze cognitieve onzuiverheden ontstaan vanuit verschillende processen die vaak moeilijk te onderscheiden zijn. Het gaat hier onder andere om vuistregels voor het verwerken van informatie (heuristieken; Kahneman, Slovic, & Tversky, 1982), beperkte capaciteit van het geheugen bij het verwerken van informatie (Simon, 1955), emotionele en morele motivaties (Pfister & Böhm, 2008), en sociale invloeden (Wang, Simons, & Brédart, 2001). Vertekeningen in oordelen kunnen voorkomen bij observatoren die anderen beoordelen en bij mensen die zichzelf beoordelen. Vertekeningen in het beeld van een observeerder over een ander wordt in de wetenschap aangeduid als *observer bias*. Vertekeningen in het beeld van iemand die zichzelf beoordeelt, worden *response biases* genoemd. Bij dit onderzoek richten we ons vooral op zelfbeoordelingen en hoe die verbeterd kunnen worden.

Om beoordelingen over anderen te kunnen objectiveren, wordt er in de praktijk veel gebruik gemaakt van (zelfbeoordelings)vragenlijsten (Verschueren & Koomen, 2007). Hoewel dit in eerste instantie een *observer bias* voorkomt, kunnen (zelf)beoordelingsvragenlijsten ook

zorgen voor vertekeningen van de werkelijkheid. Situaties waarin personen zelfstandig een beoordelingsvragenlijst invullen, zijn onderhevig aan *response biases*: de neiging om systematisch vragenlijstitems te beantwoorden op een manier die niet overeenkomt met de werkelijkheid (Paulhus, 2002). Een veel voorkomende *bias* is de neiging om het meest wenselijke of het minst extreme antwoord te kiezen, of het eens te zijn met stellingen ongeacht de inhoud daarvan om een positief beeld te creëren. Dit wordt sociale wenselijkheid genoemd (Ones, Viswesvaran, & Reiss, 1996). He en Van de Vijver (2013) beschrijven de meest voorkomende invulneigingen tijdens het maken van (zelfbeoordelings)vragenlijsten: *Acquiescence* (de neiging om het eens te zijn met items, ongeacht de inhoud daarvan), *extremity* (de neiging om de uitersten van de schalen veel te gebruiken) en *midpoint responding* (de neiging om het midden van een schaal veel te gebruiken).

Selectiebeslissingen op basis van ‘gebiaste’ gesprekken of vragenlijsten kunnen een grote impact hebben op zowel de organisatie als de persoon zelf (Risavy & Hausdorf, 2011). Een onjuiste beoordeling over een persoon kan leiden tot het aannemen van de verkeerde mensen met de enorme consequenties van dien. Een persoon die eigenlijk niet geschikt is voor de vacature selecteren kan leiden tot een situatie waarin kostbare fouten worden gemaakt door de werknemer, of tot noodzaak voor extra begeleiding en trainingen. Uiteindelijk zal een niet geschikte werknemer leiden tot zowel ontevredenheid bij de betrokkene zelf als zijn collega’s als zijn werkgevers/managers. Het aannemen van de juiste persoon leidt daarentegen tot toegenomen productiviteit en motivatie, wat zowel voor het individu als voor de organisatie voordelig is.

Een gewenste aanvulling op huidige selectiemethoden zou een objectieve beoordelingstechniek zijn, waardoor de kans op ‘gebiaste’ beoordelingen minder groot is. Het doel van dit onderzoek is het wetenschappelijk onderzoeken van een nieuwe

assessmenttechnologie die een zo objectief mogelijk beeld van een kandidaat schetst. Op basis van de respons van een kandidaat tijdens het sollicitatiegesprek wordt de spraak omgezet in tekst, waarna bepaalde voorspellende eigenschappen van de kandidaat door middel van een speciaal ontwikkeld algoritme voorspeld kunnen worden. De bijbehorende onderzoeksvraag luidt als volgt:

In hoeverre kan het geautomatiseerd beoordelen van (gesproken) tekst bestaande beoordelingen vervangen of aanvullen om werkprestatie te voorspellen?

Wetenschappelijke en praktische relevantie

Wetenschappelijk onderzoek naar in hoeverre het geautomatiseerd beoordelen van (gesproken) tekst menselijke beoordelingen kan vervangen of aanvullen om werkprestatie te voorspellen is nog niet eerder verricht. Binnen de Arbeids- en Organisationspsychologie is inmiddels vele jaren onderzoek gedaan naar de juiste manier om kandidaten te beoordelen (Eder & Harris, 1999). Wetenschappers waren geïnteresseerd in een breed scala aan onderwerpen, zoals de psychometrische functie van interviews, de manier waarop een interview wordt afgenomen (gestructureerd versus ongestructureerd), het cognitieve besluitvormingsproces van interviewers en de rol die (zelfbeoordelings)vragenlijsten kunnen spelen bij het beoordelen van kandidaten. In de wetenschap is er best wat onderzoek gedaan naar het vinden van een oplossing om zowel de kans op een *observer bias* als een *response bias* te minimaliseren. Het probleem is alleen dat deze onderzoeken vaak erg technische oplossingen opleveren die moeilijk in de praktijk zijn toe te passen. Er worden tegenwoordig weinig alternatieve meetmethoden verkend die wellicht betere eigenschappen hebben. Dit onderzoek naar geautomatiseerde beoordelingen

van gesproken tekst levert een bijdrage aan het verkennen van nieuwe meetmethoden. Omdat er geen mensen bij de beoordeling betrokken zijn en geen (zelfbeoordelings)vragenlijsten door de kandidaat worden ingevuld, treden er zo min mogelijk *response biases* op die bij zelfbeoordelingsvragenlijsten wel een rol spelen.

In de praktijk zouden geautomatiseerde beoordelingen ook relevant kunnen zijn voor organisaties in het algemeen en Human Resources afdelingen in het bijzonder. Enerzijds zou het een goede manier kunnen zijn om kosten te besparen door een computer het werk te laten doen in plaats van een (kostbare) werknemer. Anderzijds zouden geautomatiseerde beoordelingen een bijdrage kunnen leveren aan een zo objectief mogelijke beoordeling over de kandidaat te krijgen (Segrest-Purkiss et al., 2006), waardoor voorspellingen van werkprestatie kunnen worden verbeterd.

Het huidige onderzoek bestaat uit drie stappen: Allereerst wordt vastgesteld wat de beste voorspeller van werkprestatie is, zoals gemeten kan worden door middel van zelfrapportage. In het vervolg van deze inleiding wordt beschreven hoe en welke voorspeller is gekozen. Daarna wordt vastgesteld hoe deze voorspeller van werkprestatie gemeten wordt (door middel van de nieuwe assessmenttechnologie). Vervolgens wordt vastgesteld in hoeverre de predictor, gemeten met de nieuwe assessmenttechnologie, een goede voorspeller is van werkprestatie. Allereerst wordt het theoretisch kader gepresenteerd dat de basis legt voor dit onderzoek, waarna in de methodologie wordt omschreven op welke manier het onderzocht wordt. Vervolgens worden de resultaten beschreven, waarna er een discussie van de implicaties plaatsvindt en tot slot volgt een aantal voorstellen voor vervolgonderzoek.

Theoretisch kader

Persoonlijkheid en werkprestatie

Er zijn twee belangrijke factoren die prestatie op de werkvloer bepalen: can-do (capaciteit) en will-do (motivatie; Carroll, 1993; Murphy, 1996; Spearman, 1927). Dit inzicht komt veel terug in literatuur binnen de Arbeids- en Organisationspsychologie. Voorbeelden van can-do factoren zijn intelligentie en capaciteiten van lagere orde, zoals ruimtelijke waarneming, wiskundige en verbale vaardigheden en analytisch denkvermogen. Deze vaardigheden stellen iemand in staat een taak succesvol af te ronden. De will-do factor is te achterhalen op basis van iemands persoonlijkheid: Is een persoon van nature geneigd om de opgelegde taken ook zo goed mogelijk te volbrengen. Het is belangrijk om te weten dat de can-do en will-do factoren niet los van elkaar staan, maar juist een functie zijn om werkprestatie vast te stellen. Dit betekent dat iemand niet alleen over de juiste capaciteit moet beschikken, maar dat het van groot belang is dat een persoon ook deze capaciteiten inzet. Persoonlijkheid bepaalt of iemand zijn capaciteiten zal inzetten (Borman, Ilgen, & Klimoski, 2003). Als één van beide factoren laag of afwezig is, zal de prestatie onvoldoende zijn.

In het afgelopen decennium is de interesse voor persoonlijkheidsdeterminanten van werkprestatie sterk toegenomen op het gebied van onderzoek naar personeelsselectie (Borman, Ilgen, & Klimoski, 2003). Persoonlijkheidsmetingen worden ook in toenemende mate door managers en human resource professionals gebruikt om de geschiktheid van kandidaten voor verschillende posities in een organisatie te beoordelen. De groei van de populariteit van deze personeelselectietechniek komt voort uit een reeks van meta-analytisch onderzoek in het begin van de jaren '90, waarin aangetoond wordt dat de persoonlijkheid van een kandidaat een hoge mate van voorspellende validiteit voor werkprestatie heeft (Rothstein & Goffin, 2006; Schmidt

& Hunter, 1998).

Vanaf eind jaren '90 zijn wetenschappers het erover eens dat bijna alle persoonlijkheidsmetingen gecategoriseerd kunnen worden volgens het vijf-factoren model van persoonlijkheid, ook wel aangeduid als de 'Big Five' persoonlijkheidsdimensies (Goldberg, 1990; Hogan, Hogan & Roberts, 1996). Deze vijf persoonlijkheidsdimensies blijken toepasbaar en relevant te zijn voor verschillende culturen (McCrae & Costa, 1997) en worden consistent gevonden in factoranalyses van zelfbeoordelingen (Costa & McCrae, 1988). Het vijf-factoren model bestaat uit de volgende Engelse (en Nederlandse) dimensies: *Extraversion* (Extraversie), *Agreeableness* (Vriendelijkheid), *Emotional Stability* (Emotionele Stabiliteit), *Conscientiousness* (Consciëntieusheid) en *Openness to Experience* (Openheid voor Ervaringen). De afgelopen jaren is een alternatieve weergave van persoonlijkheidsstructuur veelvuldig aangetoond. Deze alternatieve structuur, die het HEXACO model wordt genoemd, bestaat uit zes persoonlijkheidsdimensies in plaats van vijf persoonlijkheidsdimensies (Ashton & Lee, 2007). De belangrijkste toevoeging van het HEXACO model is de toevoeging van de Integriteitsdimensie (Honesty-Humility). Daarnaast bevat het model ook wijzigingen in de posities van de assen van Verdraagzaamheid en Emotionaliteit; deze zijn geroteerd ten opzichte van de assen van Vriendelijkheid en Emotionele Stabiliteit in het Big Five model (De Vries, Ashton & Lee, 2009). De Vries et al. (2009) omschrijven de persoonlijkheidsdimensies als volgt: Integriteit geeft de mate van oprechtheid, rechtvaardigheid, hebzuchtvermijding en bescheidenheid aan. Extraversie geeft de mate van sociale zelfwaardering, sociale bravoure, sociabiliteit en levendigheid aan. Emotionele stabiliteit geeft de mate van angstigheid, bezorgdheid, afhankelijkheid en sentimentaliteit aan. Vriendelijkheid geeft de mate van vergevingsgezindheid, zachtaardigheid, aanpassingsbereidheid en geduld aan. Openheid voor

ervaringen geeft de mate van esthetische waardering, weetgierigheid, creativiteit en unconventionaliteit aan. Ten slotte, geeft Consciëntieusheid de mate van ordelijkheid, ijver, perfectionisme en bedachtzaamheid van een persoon aan.

Uit onderzoek blijkt dat, van alle persoonlijkheidsdimensies, Consciëntieusheid de beste voorspeller is van succesvolle prestaties op de werkvloer bij zowel managers als werknemers (bv. Barrick & Mount, 1991; Higgins, Peterson, Lee & Pihl, 2007; Salgado, 1997).

Consciëntieuze werknemers zijn over het algemeen betrouwbaarder, gemotiveerder en hardere werkers. Zij hebben ook een lagere mate van absentie en contraproductieve gedragskenmerken, zoals stelen en minder ernstige conflicten met collega's (Roberts, Jackson, Fayard, Edmonds & Meints, 2009). Lage scores op de persoonlijkheidsdimensie Consciëntieusheid gaan vaak gepaard met uitstelgedrag (Dewitt & Schouwenburg, 2002). Daarnaast is Consciëntieusheid de enige persoonlijkheidsdimensie die correleert met werkprestatie bij alle soorten banen (Mount, Barrick & Stewart, 1998). Hurtz en Donovan (2000) presenteren in hun onderzoek de resultaten van een meta-analyse over beroepen en werkprestatie. Deze analyses zijn gebaseerd op 35-45 correlaties en 5,525-8,083 sollicitanten en werknemers (De exacte correlaties en het aantal sollicitanten en werknemers verschilt per persoonlijkheidsdimensie). De resultaten bevestigden het eerdere onderzoek van Barrick en Mount (1991) en Salgado (1997) dat Consciëntieusheid de hoogste predictieve validiteit heeft van alle persoonlijkheidsdimensies.

Persoonlijkheid meten

In organisaties wordt persoonlijkheid in de meeste gevallen gemeten door zelfbeoordelingsvragenlijsten om beoordelingen over anderen te kunnen objectiveren (Dodorico-McDonald, 2008). De belangrijkste redenen voor het gebruik van zelfbeoordelingsvragenlijsten

zijn het gemak, de lage kosten en de goede betrouwbaarheid en validiteit (Schmidt & Hunter, 1998). In de inleiding is beschreven dat er verschillende *biases* zijn bij interviews en vragenlijsten. Persoonlijkheidsvragenlijsten zijn ook aan dergelijke *biases* onderhevig. Hoewel dit in eerste instantie een *observer bias* voorkomt, kunnen zelfbeoordelingsvragenlijsten ook zorgen voor vertekeningen van de werkelijkheid door de zogenoemde *response bias* (Paulhus, 2002). Bovendien maken kandidaten gebruik van allerlei invulstrategieën die kunnen zorgen voor vertekeningen in het beeld van de werkelijkheid. Kleinmann et al. (2011) hebben onderzoek gedaan naar de invloed van een kandidaats vermogen om na te gaan welke prestatiecriteria beoordeeld worden tijdens een assessment, oftewel de *Ability to Identify Criteria* (ATIC). De aanname is dat kandidaten tijdens een selectieprocedure zich afvragen welke criteria beoordeeld zullen worden en dat deze assumpties vervolgens leiden tot bepaald gedrag om deze criteria te beïnvloeden. Deze invulstrategie heeft een positieve invloed op de criteriumvaliditeit van de meting, maar er ontstaat ook een vertekening in het beeld van de werkelijkheid doordat de voorspelling niet alleen op basis van persoonlijkheid wordt gedaan.

Een andere mogelijke manier waarop individuen antwoorden tijdens het maken van zelfbeoordelingsvragenlijsten is het zogenoemde *reference-group effect* (RGE; Credé, Bashshur, & Niehorster, 2010). De RGE weerspiegelt de invloed van het vergelijken van anderen (referentiegroep) tijdens het invullen van zelfbeoordelingsvragenlijsten. Zo kan de respondent een andere score op een persoonlijkheidsdimensie ervaren in vergelijking met zijn familie, vrienden of collega's. Afhankelijk van de keuze van de referentiegroep komt de score op een persoonlijkheidsdimensie tot stand, welke kan afwijken van iemands werkelijke score op een persoonlijkheidsdimensie.

Nu vastgesteld is dat het meten van persoonlijkheid aan de hand van persoonlijkheidsvragenlijsten niet zonder beperkingen is, rijst de vraag: Zou er een instrument ontwikkeld kunnen worden dat beoordelingen kan geven over iemands persoonlijkheid met zo min mogelijk vertekeningen?

Nieuwe assessmenttechnologie

De woorden die we in het dagelijks leven gebruiken weerspiegelen wie we als persoon zijn en in welke sociale relaties wij ons bevinden. Het gebruiken van woorden is een manier voor mensen om hun gedachten en emoties begrijpelijk te maken voor anderen (Tausczik & Pennebaker, 2010). Woorden geven aan waar wij als mens aandacht aan besteden, waar wij aan denken, wat wij proberen te vermijden, hoe wij ons voelen en hoe wij onze eigen wereld organiseren en analyseren. Dat mensen verschillen in de manier waarop zij spreken en de keuze van woorden die ze daarbij maken, mag geen verrassing zijn. Zelfs als de inhoud van de boodschap hetzelfde is, gebruiken mensen verschillende stijlen om zichzelf te uiten. Enkele onderzoeken hebben aangetoond dat het verschil in woordgebruik iets zegt over de persoonlijkheid van mensen (Fast & Funder, 2008; Pennebaker & King, 1999).

Om op een efficiënte en effectieve manier de verschillende emotionele en cognitieve structuren en processen die voortkomen uit woord en geschrift te bestuderen, is er vanaf 1993 gewerkt aan een tekst analyse applicatie genaamd *Linguistic Inquiry and Word Count* (LIWC) (Pennebaker, 1993). Momenteel wordt er in wetenschappelijk onderzoek vooral gebruik gemaakt van LIWC2007 (Pennebaker, Booth, & Francis, 2007). LIWC is een tekstanalyse softwareprogramma dat de mate berekent waarin mensen woorden gebruiken in verschillende categorieën om deze vervolgens te linken aan psychologische dimensies, zoals

persoonlijkheidsdimensies, om gedrag te kunnen voorspellen. Eén van de eerste validiteitstests die het LIWC systeem onderging, was uitgevoerd door Pennebaker en Francis (1996) als onderdeel van een experiment waarbij eerstejaarsstudenten moesten schrijven over hun ervaringen tijdens de eerste dagen van hun nieuwe studieperiode. 72 studenten Psychologie kwamen als groep bijeen en moesten drie opeenvolgende dagen schrijven over hun toegewezen onderwerpen. Participanten in de experimentele groep ($N=35$) werden geïnstrueerd om te schrijven over hun diepste gedachtes en gevoelens met betrekking tot de eerste ervaringen van de studie. De participanten in de controle groep ($N=37$) werden gevraagd om een onderwerp naar keuze te beschrijven op een niet-emotionele manier. Nadat deze fase van de studie was afgerond, beoordeelden vier assessoren de essays van de participanten op verschillende emotionele en cognitieve dimensies die overeenkomen met geselecteerde *LIWC Dictionary Scales*. Vervolgens werden de output van het LIWC tekstanalyse softwareprogramma en de beoordelingen van de assessoren op basis van een Pearson correlatieanalyse met elkaar vergeleken om de validiteit te testen. Uit de resultaten van de analyses blijkt dat de LIWC schalen en de beoordelingen van assessoren sterk gecorreleerd zijn bij sociale processen ($r= .87$), affectieve processen ($r= .41$) en cognitieve processen ($r= .44$).

Dit gegeven bood wetenschappers de juiste handvatten om aan de hand van LIWC onderzoek te doen naar de mogelijkheid om persoonlijkheid te beschrijven op basis van woordgebruik. Onderzoeken hiernaar varieerden van het analyseren van e-mails (Oberlander & Gill, 2006), geschreven persoonlijke verhalen (Hirsch & Peterson, 2009), het opnemen van gesprekken (Mehl, Gosling, & Pennebaker, 2006; Mehl & Pennebaker, 2003), tot persoonlijke blogs (Yarkoni, 2010). Ook deze onderzoeken bevestigden dat de relatieve frequentie van het gebruik van bepaalde woorden afhangt van iemands persoonlijkheid. Echter, de focus van deze

onderzoeken lag vooral op de persoonlijkheidsdimensie Extraversie en Emotionaliteit. Er is weinig tot geen aandacht besteed aan het analyseren van (gesproken) tekst in combinatie met de persoonlijkheidsdimensie Consciëntieusheid, terwijl dit in een werkcontext de meest relevante persoonlijkheidsdimensie is.

Het geautomatiseerd beoordelen van alle persoonlijkheidsdimensies (Big Five) op basis van een stuk tekst is een aantal jaar geleden met behulp van LIWC grootschalig door wetenschappers onderzocht (Mairesse, Walker, Mehl, & Moore, 2007). Er is onderzoek verricht naar het gebruik van *Linguistic Cues* voor het automatisch herkennen van persoonlijkheid in conversaties en tekst. De uitkomsten van dit onderzoek toonden een Pearson's correlatiecoëfficiënt, tussen de verwachte scores op de persoonlijkheidsdimensies en de werkelijke scores (zelf en ander beoordeling) op de persoonlijkheidsdimensies, tussen de 0.44 en 0.54.

Alhoewel de LIWC analyses wetenschappelijk bewezen bruikbaar zijn voor het detecteren van persoonlijkheid, zijn de resultaten van de LIWC analyses nog steeds gebaseerd op het uitsluitend tellen van woorden, in plaats van de context waarin de woorden zijn gebruikt (Zijlstra et al., 2005). Veel woorden zijn, zodra ze losstaan van hun context, ambigu. Elke LIWC categorie waar een persoonlijkheidsdimensie aan gelinkt wordt, bevat een verscheidenheid aan woorden die, ondanks dat ze statistisch gezien bij elkaar horen, op verschillende manieren geïnterpreteerd kunnen worden. Omdat geschreven tekst op meerdere manieren tegelijk geanalyseerd en geïnterpreteerd kan worden, zoals een woord, zin, paragraaf of pagina, zal door de beperking van het LIWC programma veel bruikbare data over iemands persoonlijkheid verloren gaan tijdens het simpel tellen van woorden. Hirsch en Peterson (2009) geven aan dat analysetechnieken, waarbij semantische structuren van hogere orde worden meegenomen zeer

gewenst zijn voor vervolgonderzoek.

Het bedrijf NeedForward heeft speciaal voor dit onderzoek het sentiment analyse algoritme product, genaamd Sentimentics, beschikbaar gesteld (Sentimentics, 2014). Dit algoritme is in staat om op basis van een stuk tekst geautomatiseerd verschillende eigenschappen en karaktertrekken te detecteren en te meten op een ratio schaal. Het algoritme kan metingen verrichten in termen van sentiment en, specifiek relevant voor dit onderzoek, in termen van persoonlijkheidstrekken. Ook worden conflictwaarden teruggegeven voor elke dimensie van analyse (voor dit onderzoek de HEXACO persoonlijkheidsdimensies). Daarnaast worden grammaticale structuren als input variabele meegenomen voor analyse. Een verbetering ten opzichte van het LIWC-programma is dat deze nieuwe assessmenttechnologie persoonlijkheid op twee verschillende manieren kan detecteren, zodat het tegemoet komt aan de beperkingen van het LIWC-programma en de suggestie voor vervolgonderzoek van Hirsch en Peterson (2009). Ten eerste maakt het algoritme gebruik van versterkende en verzwakkende woorden. Dit is een sterke verbetering ten opzichte van het LIWC, omdat bij het LIWC alleen de woorden worden geteld en deze direct worden gelinkt aan een persoonlijkheidsdimensie. Het algoritme herkent sentiment uit tekst door verschillende taalkundige en grammaticale elementen uit de tekst te detecteren en combinaties van deze te analyseren in termen van sentiment. Daarnaast is het mogelijk om verschillende *keywords* in te voeren, waarbij elk woord van een stuk tekst positief of negatief wordt gelinkt aan een karaktertrek. Het algoritme kan de mate van ambiguïteit meten, waardoor er meer nuance wordt gelegd dan het LIWC-programma. Tenslotte wordt er gelet op interpunctie, waardoor het algoritme herkent welke woorden er samenhangen. Het algoritme kan woorden lichter, zwaarder of omgekeerd wegen, afhankelijk van andere woorden waaraan het is verbonden.

Op basis van bovenstaande informatie zijn twee punten vastgesteld: 1) De persoonlijkheidsdimensie Consciëntieusheid is de beste voorspeller van werkprestatie, gemeten met zelfrapportage (Salgado, 1997; Higgins et al., 2007) en 2) persoonlijkheid is te achterhalen op grond van woordgebruik in een stuk tekst. Om deze twee redenen is gekozen om persoonlijkheid - en specifiek Consciëntieusheid - te gebruiken als voorspeller in dit onderzoek. Omdat er geen mensen bij de beoordeling betrokken zijn, krijgt men waarschijnlijk een zo objectief mogelijk beeld van de kandidaat. Vervolgens kan er met een objectieve maat van Consciëntieusheid algemene werkprestatie voorspeld worden.

Als eerste moet het algoritme op basis van het woordgebruik van een kandidaat de mate van Consciëntieusheid kunnen voorspellen. Bij dit onderzoek beperken we ons tot het meten van Consciëntieusheid, omdat deze persoonlijkheidsdimensie de beste voorspeller is van werkprestatie. We meten Consciëntieusheid op twee manieren: zelfbeoordeling en geautomatiseerde beoordeling. Gebaseerd op eerdere LIWC onderzoeken waarin de persoonlijkheid van de deelnemers geautomatiseerd beoordeeld werd (Mairesse, Walker, Mehl, Moore, 2007), wordt verwacht dat de samenhang sterker zal zijn dan een Pearson's correlatiecoëfficiënt tussen de 0.44 en 0.54. Het speciaal ontwikkelde algoritme is namelijk geavanceerder dan het LIWC dat in voorgaande onderzoeken werd gebruikt, waardoor een sterke samenhang verwacht wordt. Op basis van de volgende hypothese wordt de constructvaliditeit getoetst van de geautomatiseerde meting:

H1: De geautomatiseerde beoordelingen van Consciëntieusheid komen gemiddeld tot sterk overeen met de Consciëntieusheid zelfbeoordelingen

Nadat de constructvaliditeit van de geautomatiseerde beoordelingen van Consciëntieusheid en de Consciëntieusheidsbeoordelingen door middel van persoonlijkheidsvragenlijsten vastgesteld is, moet worden nagegaan in hoeverre de verschillende constructen de mate van werkprestatie kunnen voorspellen. Op basis van de volgende hypothesen wordt de predictieve validiteit getoetst:

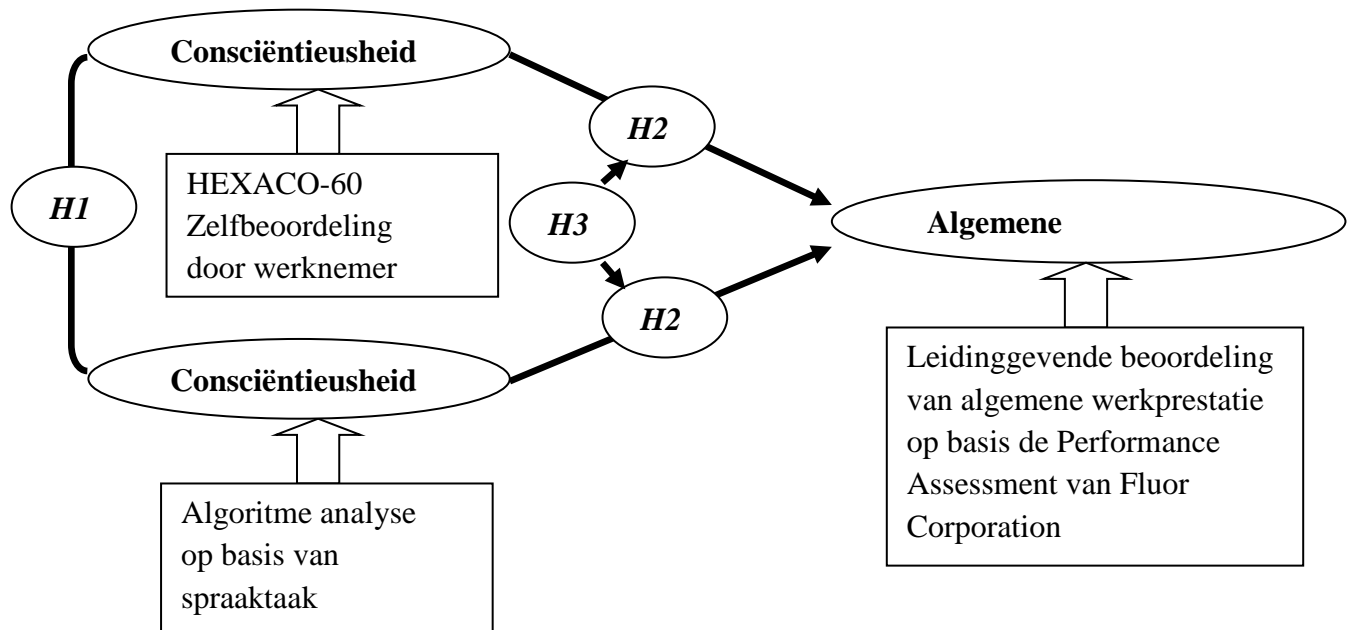
H2a: Consciëntieusheid zelfbeoordelingen voorspellen de mate van werkprestatie.

H2b: Geautomatiseerde Consciëntieusheidsbeoordelingen voorspellen de mate van werkprestatie.

De verwachting is dat Consciëntieusheid zelfbeoordeling de mate van werkprestatie beter kan voorspellen dan de geautomatiseerde Consciëntieusheidsbeoordeling. Verklaring hiervoor is dat de HEXACO-PI vragenlijst, waarmee de zelfbeoordeling tot stand komt, in eerdere onderzoeken al valide en betrouwbaar is bevonden (De Vries, 2006). Het ontwikkelde algoritme wordt voor het eerst aan een serieuze test onderworpen, waardoor de kans op minder valide en betrouwbare resultaten groter is.

H3: Consciëntieusheid zelfbeoordelingen hebben een grotere voorspellende waarde voor werkprestatie dan geautomatiseerde Consciëntieusheidsbeoordelingen.

Kortom, eerst wordt de constructvaliditeit van de geautomatiseerde meting onderzocht. Vervolgens wordt ook de predictieve validiteit onderzocht. Alle hypothesen worden visueel weergegeven in figuur 1.



Figuur 1. Het conceptueel model

Onderzoeksmethode

Onderzoeksdesign

Dit onderzoek had als doel te onderzoeken in hoeverre automatische beoordelingen gestructureerde vragenlijsten kunnen aanvullen of vervangen. Er werd gekeken naar de verbanden tussen de twee onafhankelijke variabelen: Consciëntieusheid zelfbeoordeling en geautomatiseerde Consciëntieusheidsbeoordeling. Deze maten van Consciëntieusheid worden vervolgens gerelateerd aan de afhankelijke variabele Algemene werkprestatie. Het onderzoeksdesign betreft een within-subject design, elke deelnemer werd getest onder dezelfde voorwaarden en onderworpen aan meerdere tests. Daarnaast is het onderzoek cross-sectioneel uitgevoerd. De dataverzameling van de voorspellers en uitkomsten vindt plaats op één moment in de tijd en bij één groep. Het voordeel van dit type onderzoek is dat data relatief snel verzameld kan worden en in vergelijking met longitudinaal onderzoek is er geen risico op het uitvallen van participanten.

Procedure

Alle participanten zijn via Lotus Notes (e-mail client) benaderd met het verzoek om deel te nemen aan dit onderzoek. Op basis van het personeelsbestand is er uiteindelijk naar 461 werknemers een uitnodiging gestuurd. Het onderwerp en doel van het onderzoek werd in de uitnodiging omschreven, plus de taken die de respondenten moeten uitvoeren. In het eerste bericht werd er aan de participanten gevraagd om de HEXACO-60 persoonlijkheidsvragenlijst in te vullen. Er werd voor de zelfbeoordeling gebruik gemaakt van een digitale enquête, verspreid via Qualtrics (Online Survey Software). Totaal hebben 125 werknemers de HEXACO-60 ingevuld. Daarna werd naar de medewerkers die hadden deelgenomen aan de

persoonlijkheidsvragenlijst een tweede bericht gestuurd met het verzoek om deel te nemen aan de spraaktaak. Er werd benadrukt dat het onderzoek vrijblijvend en strikt vertrouwelijk was. Uiteindelijk hebben 103 medewerkers ook de spraaktaak gemaakt.

Participanten

De betrokken onderzoekspopulatie is samengesteld uit werknemers van een groot ingenieursbureau in Haarlem en Bergen op Zoom, waar opgeteld meer dan 800 mensen werken. Er waren twee belangrijke exclusiecriteria: De eerste was dat er een beoordeling van de algemene werkprestatie van de participant moest zijn om de predictieve validiteit van de spraaktaak te kunnen toetsen. Dit sloot echter werknemers zonder beoordelingsgesprekken niet uit, omdat op basis van het invullen van de vragenlijst en het uitvoeren van de spraaktaak wel de constructvaliditeit aangetoond kon worden. Daarnaast is het algoritme alleen in staat om Nederlandstalige teksten te analyseren. Werknemers die niet de Nederlandse taal machtig waren, konden niet aan het onderzoek deelnemen. In totaal hebben 125 medewerkers van de 461 (27,1% respons) aan het onderzoek meegewerkt, waarvan er 103 medewerkers (22,3%) zowel aan de vragenlijst als aan de spraaktaak hebben deelgenomen. Uiteindelijk waren de data van 90 van de 103 respondenten bruikbaar om alle hypothesen mee te testen. Bij de overige 13 respondenten kon er geen goede meting door middel van de spraaktaak uitgevoerd worden, doordat de deelnemers te korte antwoorden hebben gegeven.

Van de participanten zijn er 106 man en 19 vrouw. De participanten hebben een gemiddelde leeftijd van afgerond 40 jaar ($M= 39,62$, $SD = 12,02$) met een gemiddelde werkervaring van gemiddeld 14 jaar ($M=13,75$, $SD = 12,35$). De hoogste afgeronde of huidige opleiding van de deelnemers was Hoger beroepsonderwijs (46,4%),

Wetenschappelijk onderwijs (28,8%) of Middelbaar beroepsonderwijs (19,2%).

Meetinstrumenten

Consciëntieusheid (Zelfbeoordeling)

De werknemers maakten tijdens het onderzoek gebruik van de persoonlijkheidsvragenlijst HEXACO-PI- 60 om zichzelf te beoordelen (Ashton & Lee, 2009). De HEXACO-PI-60 is een korte persoonlijkheidsvragenlijst, gebaseerd op de volledige persoonlijkheidsvragenlijst HEXACO Personality Inventory–Revised (HEXACO-PI-R). Deze vragenlijsten meten zes persoonlijkheidsdimensies: Integriteit, Emotionaliteit, Extraversie, Verdraagzaamheid, Consciëntieusheid en Openheid voor ervaringen (zie bijlage 2 – HEXACO-PI-60). Bij de HEXACO- PI-60 zijn per persoonlijkheidsdimensie 10 items geselecteerd van de volledige HEXACO PI-R (Ashton & Lee, 2009). De items worden beantwoord aan de hand van een 5-punts Likertschaal met antwoord categorieën van 1= *helemaal mee oneens* tot 5= *helemaal mee eens*. Alleen de scores op de persoonlijkheidsdimensie Consciëntieusheid werden in het huidige onderzoek meegenomen. De reden dat de volledige vragenlijst is gebruikt, is dat niet overduidelijk mocht zijn wat er gemeten werd tijdens dit onderzoek. De schalen van de HEXACO-PI-60 tonen hoge niveaus van interne consistentie en betrouwbaarheid. Bij de HEXACO-PI-60 waren de alfa betrouwbaarheden in eerder onderzoek (De Vries, 2006) .77 voor Integriteit, .75 voor Emotionaliteit, .79 voor Extraversie, .71 voor Verdraagzaamheid, .71 voor Consciëntieusheid en .77 voor Openheid voor Ervaringen. De HEXACO-PI-60 schalen vertoonden over het algemeen lage onderlinge correlaties (alle < .30) en hoge convergente correlaties en lage discriminante correlaties met de factorschalen van de volledige HEXACO-PI-R (alle convergente correlaties > .87; alle discriminante correlaties < .33). Ook de convergente

correlaties tussen de zelfbeoordeling en de ander beoordeling van de HEXACO-60 waren hoog, met een gemiddelde van boven de .50. Ten slotte, waren de uitkomsten van de HEXACO-PI-60 over het algemeen vergelijkbaar met de langere versies van de HEXACO-PI-R (Ashton & Lee, 2009). De HEXACO-PI-60 is door zijn korte lengte en acceptabele psychometrische eigenschappen ideaal voor onderzoek naar persoonlijkheid waarbij de tijd beperkt is.

De alfa betrouwbaarheden van de verschillende schalen van de HEXACO-PI-60 gebaseerd op data voor dit onderzoek (Integriteit= .69, Emotionaliteit= .75, Extraversie= .73, Verdraagzaamheid= .68, Consciëntieusheid= .76, en Openheid voor Ervaringen= .73), kwamen in grote lijnen overeen met de gevonden alfa betrouwbaarheden in eerdere onderzoeken. Opvallend is dat de alfa betrouwbaarheden van Integriteit en Verdraagzaamheid onder de algemeen aangenomen ondergrens van .70 zijn. Bij de persoonlijkheidsdimensie Integriteit heeft vraag 6 “Ik zou niet vleien om op het werk opslag of promotie te krijgen, zelfs al zou het succes hebben.” een negatieve invloed op de betrouwbaarheid. Als dit item verwijderd zou worden, dan komt de betrouwbaarheid op .73 en voldoet het aan de algemeen aangenomen ondergrens van .70.

Consciëntieusheid (Geautomatiseerde beoordeling)

De persoonlijkheidsdimensie Consciëntieusheid werd ook geoperationaliseerd door de nieuwe assessmenttechnologie, het algoritme. Om tekst te verzamelen die ingevoerd kon worden in het algoritme, moesten de participanten een aantal vragen beantwoorden. Deze vragen zijn zo veel mogelijk gebaseerd op bestaand materiaal. Specifiek de facetten Ordelijkheid en IJver van de persoonlijkheidsdimensie Consciëntieusheid zijn gebaseerd op vragen die ook bij de organisatie, waar dit onderzoek is uitgevoerd, tijdens sollicitatiegesprekken worden gevraagd.

Resultaatgerichtheid en verantwoordelijkheid zijn twee van de 12 *Core Skills* waarop werknemers in de organisatie worden beoordeeld. De *Core Skill* Resultaatgerichtheid (*Drive for Results*) wordt omschreven als “*Has clearly articulated goals and is focused on achieving them consistently. Is responsive and timely. Personally gets results and helps others by supporting, encouraging and coaching them as needed*”. Verantwoordelijkheid (*Accountability*) wordt omschreven als “*Sets high goals and standards personally and with team members. Can be counted on to deliver what has been agreed to. Plans, prioritizes and executes work based on clearly defined outcomes and expectations. Is clear about the results desired.*” Deze *Core Skills* liggen het dichtst bij de persoonlijkheidsdimensie Consciëntieusheid, waardoor de verwachting is dat Consciëntieusheid werkprestatie (en dus zeker deze twee *Core Skills*) voorspellen. Hier zijn een aantal vragen over Perfectionisme en Bedachtzaamheid aan toegevoegd om alle vier de facetten (Ordelijkheid, IJver, Perfectionisme en Bedachtzaamheid) van Consciëntieusheid te kunnen meten. Tot slot werd er een algemene Consciëntieusheidsvraag gesteld om het aantal hits op persoonlijkheidsdimensie Consciëntieusheid te vergroten. Kandidaten kregen per vraag twee minuten om antwoord te geven. De response van de kandidaten is woord voor woord uitgeschreven en vervolgens is deze tekst aangeboden aan het algoritme dat de score op de verschillende persoonlijkheidsdimensies berekent op basis van ingevoerde Consciëntieusheidsadjectieven.

Het algoritme, genaamd Sentimentics, is een product van NeedForward dat de mogelijkheid biedt om sentiment te detecteren op basis van tekst met grote precisie. Het algoritme is in staat om tekst te analyseren op verschillende semantische niveaus. De precisie van het algoritme is door meerdere interne studies op basis van twitter *corpusses* aangetoond met waarden van 0.92 en 0.94 voor het classificeren van tweets in de categorieën positief, negatief en

neutraal (Sentimentics, 2014). Wanneer het algoritme tekst analyseert dan let het op een aantal zaken om de scores op de persoonlijkheidsdimensies te kunnen bepalen. Belangrijk is dat het algoritme de lijst van 551 bijvoeglijk naamwoorden van het lexicale onderzoek naar de HEXACO gebruikt. De Consciëntieusheidsadjectieven zijn verkregen op basis van de lading op de Consciëntieusheidsfactor in Nederlands persoonlijkheidsonderzoek (De Vries, 2006). Al deze woorden hebben een factorlading van -1 tot en met 1 op alle zes de persoonlijkheidsdimensies. De spraaktaakvragen waren er op gericht om zo veel mogelijk hits te creëren op Consciëntieusheidsgerelateerde woorden. Het algoritme scant de tekst voor deze ‘kernwoorden’, eventuele varianten op deze kernwoorden pikt het algoritme ook op. Wanneer het algoritme een kernwoord vindt in de tekst dan kijkt het ook naar de context van de zin: Versterkende (bv. heel erg), verzwakkende (bv. een beetje) omgedraaide (bv. niet) woorden beïnvloeden het gewicht van het kernwoord. Het is dus van groot belang voor het algoritme dat er ‘hits’ worden gescoord op de lijst met woorden. Alle woorden worden vervolgens bij elkaar gecombineerd in persoonlijkheidsdimensie scores. Bij 13 kandidaten waren er te weinig hits gescoord om een zinnige Consciëntieusheidscores te berekenen.

Om de kans op het aantal treffers door het algoritme te vergroten, werd er in de instructie aangegeven dat de respondent tijdens het beantwoorden van de vragen zoveel mogelijk gebruik moet maken van bijvoeglijk naamwoorden. Er werden enkele voorbeelden van mogelijke bijvoeglijk naamwoorden gegeven (*zorgvuldig, ordelijk, impulsief, flexibel, laks, gemakzuchtig, ambitieus, nonchalant, verantwoordelijk, onverschillig*) en daarnaast werden de respondenten aangemoedigd om eigen bijvoeglijk naamwoorden te gebruiken om zichzelf te omschrijven (zie bijlage 3 - Spraaktaak).

Op basis van de alfa betrouwbaarheden kan gezegd worden dat de kwaliteit van het algoritme tot dusver ondermaats is (Integriteit= .38, Emotionaliteit= .08, Extraversie= -.03, Verdraagzaamheid= .38, Consciëntieusheid= .33, Openheid voor Ervaringen= .26). De betrouwbaarheid is uitgerekend op basis van de samenhang tussen de scores op de vijf vragen. Een mogelijke verklaring voor de lage betrouwbaarheden zou kunnen zijn dat de vragen niet duidelijk genoeg waren bij de respondenten, waardoor de vragen te veel open waren voor interpretatie. Er is bewust gekozen voor minder letterlijke vragen, omdat dit over het algemeen tot meer persoonlijkheid gerelateerde antwoorden leidt (Van Iddekinge, Raymark, & Roth, 2005). Op basis van de feedback van de respondenten komt naar voren dat veel respondenten de vragen niet duidelijk genoeg vonden en dat zij niet precies wisten hoe hun antwoord eruit moest zien. Respondenten kwamen vaak met de suggestie om de vragen te stellen aan de hand van een bepaalde case, zodat zij zich makkelijker in de situatie kunnen verplaatsen. Een andere verklaring voor de lage betrouwbaarheden kan het kleine aantal metingen ($N= 90$) zijn. De hoeveelheid metingen heeft namelijk een groot effect op de betrouwbaarheid (Baarda, 2014).

Daarnaast kan er op basis van de factoranalyse geconcludeerd worden dat de interne structuur van de spraaktaak zwak is (zie bijlage 1 - Factoranalyses). Er zijn geen herkenbare onderliggende factoren te herleiden op basis van de data uit de factoranalyse. Bovendien blijkt dat vraag 4 “U krijgt een opdracht of project toegewezen waarbij u zowel aan de wensen van de klant, als de wensen van uw collega’s en leidinggevenden moet voldoen. Hoe gaat u hiermee om?” niet hetgeen te meten wat gemeten zou moeten worden. De Consciëntieusheidscore van vraag 4 is zelfs negatief gecorreleerd aan de andere metingen van Consciëntieusheid. In de verdere analyses zal spraaktaak 4 niet meegenomen worden en zal het gemiddelde van de andere vragen worden gehanteerd als geautomatiseerde Consciëntieusheidscore. Een mogelijke

verklaring kan liggen in het feit dat de *Confliction Value*, oftewel de mate van het gebruik van zowel positief als negatief geladen bijvoeglijknaamwoorden, erg hoog ligt bij deze vraag in vergelijking met de andere vragen (zie bijlage 5 – Confliction Value).

Ondanks de matige kwaliteit van deze operationalisatie van Consciëntieusheid, is het toch zinvol om analyses uit te voeren. Een lage betrouwbaarheid betekent namelijk niet dat de validiteit hier noodzakelijk onder te lijden heeft (Burisch, 1997; Paunonen & Jackson, 1985). Kortom, alhoewel de alfa betrouwbaarheden vaak zwak zijn in kleine persoonlijkheidsschalen, zijn test-retest consistentie, beoordelaar consistentie en validiteit over het algemeen van acceptabele niveaus (De Vries, 2013). De geautomatiseerde metingen komen overeen met dergelijke kleine persoonlijkheidsschalen, omdat er slechts vijf vragen worden gesteld om de persoonlijkheidsdimensie Consciëntieusheid te meten.

Algemene werkprestatie

Het criteriumdomein Algemene werkprestatie wordt geoperationaliseerd door het beoordelingsformulier van Fluor (zie bijlage 4 - Beoordelingsformulier). De werkprestatie van de participanten wordt vastgesteld door de managers die reeds een beoordelingsformulier hebben ingevuld aan de hand van de *Performance Assessment* van Fluor. De totaalscore wordt voor de beoordeling van de algemene werkprestatie gebruikt.

Op grond van de factoranalyse (zie bijlage 1 - Factoranalyses) die is uitgevoerd op basis van de *Core Values* en de *Performance Factors* van de *Performance Assessments* (Beoordeling van de Algemene werkprestatie) van de kandidaten, kan gezegd worden dat er twee onderliggende factoren onderscheiden kunnen worden. Het blijkt dat de onderdelen *Excellence*, *Teamwork*, *Knowledge of Work*, *Quality of Work* en *Quantity of Work* samen een factor vormen.

In het vervolg van het onderzoek zal deze factor *Hard Performance* worden genoemd. Daarnaast vormen *Health, Safety & Environmental* en *Integrity* een tweede factor, welke *Soft Performance* genoemd zal worden.

Resultaten

Hypothese 1

De correlaties tussen de variabelen geslacht, leeftijd, de HEXACO persoonlijkheidsdimensies, Consciëntieusheid spraaktaal en drie vormen van werkprestatie staan vermeld in tabel 1. Deze correlaties worden gebruikt om aan te kunnen tonen of twee variabelen met elkaar samenhangen. Op deze manier kan gecontroleerd worden of er sprake is van constructvaliditeit door middel van de correlatie tussen de HEXACO Consciëntieusheid en de spraaktaal Consciëntieusheid. Daarnaast staan de verschillende betrouwbaarheden van de variabelen in de diagonaal van de matrix vermeld.

In tabel 2 staan de resultaten van het gemiddelde en de standaarddeviatie van alle spraaktaal persoonlijkheidsdimensies beschreven. De gemiddelde scores van alle spraaktaken op de persoonlijkheidsdimensies zijn Integriteit= 0.19, Emotionaliteit= 0.10, Extraversie= 0.16, Verdraagzaamheid= 0.21, Consciëntieusheid= 0,17 en Openheid voor Ervaringen= 0.11. In vergelijking met de gemiddelde scores op de HEXACO-PI vragenlijst (Integriteit= 3.73, Emotionaliteit= 2.60, Extraversie= 3.74, Verdraagzaamheid= 3.20, Consciëntieusheid= 3.71 en Openheid voor Ervaringen= 3.37) is er een vergelijkbaar patroon te herkennen, waarbij de scores op de persoonlijkheidsdimensie Emotionaliteit laag zijn en de scores op de persoonlijkheidsdimensies Integriteit en Verdraagzaamheid hoog zijn.

Uit de correlaties in tabel 1 blijkt dat er een significant verband is tussen Consciëntieusheid gemeten met de HEXACO-PI-60 (HEXACO Consciëntieusheid) en Consciëntieusheid gemeten met de spraaktaal (spraaktaal Consciëntieusheid) ($r = .27, p < .01, N = 90$). De correlatie tussen het gemiddelde van de losse Spraaktaken en Consciëntieusheid heeft een iets verband ($r = .29, p < .01, N = 90$). Als vraag 4 van de spraaktaal wordt geschrapt, dan

ontstaat het sterkste verband tussen enerzijds HEXACO Consciëntieusheid en anderzijds de spraaktaak Consciëntieusheid ($r = .33, p < .01, N = 90$). Er kan dus geconcludeerd worden dat de scores op de HEXACO vragenlijst en de spraaktaak met elkaar samen hangen. Een hogere score bij de één leidt doorgaans tot een hogere score bij de ander. Op basis van algemene vuistregels voor de interpretatie van correlatiecoëfficiënten binnen de assessmentpsychologie is een correlatie van .10 klein, .30 gemiddeld en .50 groot (Cohen, 1988). De gevonden correlaties in dit onderzoek zijn op basis van deze algemene vuistregels dus ongeveer klein tot gemiddeld. Beide variabelen verklaren 10,9% ($r^2 = 0.109$) bij elkaar variantie. Hypothese 1, waarin verwacht wordt dat geautomatiseerde beoordelingen van Consciëntieusheid overeenkomen met de Consciëntieusheidzelfbeoordelingen, kan gedeeltelijk worden aangenomen. Er moet wel een kanttekening geplaatst worden bij het percentage verklaarde variantie.

Hypothese 2

Allereerst is er een significant verband tussen HEXACO Consciëntieusheid en Algemene werkprestatie ($r = .37, p < .01$). Het verband tussen deze twee variabelen wordt nog sterker wanneer *Hard Performance* in plaats van *Mean Performance* in de analyse wordt meegenomen ($r = .39, p < .01$). Het percentage verklaarde variantie is respectievelijk 13,7% en 15,2%. Op basis van de algemene vuistregels is ook dit verband zwak. Op basis van bovenstaande resultaten kan de hypothese 2a, waarin verwacht wordt dat Consciëntieusheid zelfbeoordelingen de mate van werkprestatie voorspelen, worden aangenomen.

Daarnaast is er een significant verband tussen spraaktaak Consciëntieusheid en Algemene werkprestatie ($r = .26, p < .05$). Het verband tussen deze twee variabelen wordt niet sterker wanneer *Hard Performance* in plaats van *Mean Performance* in de analyse wordt meegenomen

($r = .24, p < .05$). Hypothese 2b, waarin verwacht wordt dat geautomatiseerde Consciëntieusheidsbeoordelingen de mate van werkprestatie voorspellen, kan ook worden aangenomen.

Hypothese 3

Tenslotte zijn in de meervoudige regressieanalyse (zie tabel 3) de verschillende variabelen in combinatie met de afhankelijke variabele Algemene werkprestatie weergegeven om de causaliteit tussen de variabelen te onderzoeken. Voor het voorspellen van Algemene werkprestatie (*Mean Performance*) is een stapsgewijze meervoudige regressieanalyse uitgevoerd, waarbij *Mean Performance* eerst voorspeld is op basis van Leeftijd en Geslacht. Dat leverde een multiple correlatie op van 0.16 ($p = .34$). In stap 2 is daar HEXACO Consciëntieusheid aan toegevoegd. Dat leverde een multiple correlatie van 0.42 ($\Delta R = .26, p = .00$) op. Het toevoegen van spraaktaak Consciëntieusheid als derde variabele leverde een multiple correlatie van 0.45 ($\Delta R = .03, p = 0.10$). Voor de voorspelling van de Algemene werkprestatie lijkt dus te kunnen worden volstaan met Consciëntieusheid zelfbeoordeling (HEXACO). De toegevoegde voorspellende waarde van de spraaktaak is zeer gering en bovendien is de toegevoegde voorspellende waarde niet significant. De correlatie tussen Consciëntieusheid spraaktaak en Algemene werkprestatie is op zich significant. Echter, het heeft geen significant toegevoegde waarde bovenop de HEXACO meting. Men kan de predictieve validiteit niet verhogen door een spraaktaak toe te voegen aan de beoordeling. Kortom, de spraaktaak heeft geen incrementele validiteit.

Voor het voorspellen van *Hard Performance* is ook een stapsgewijze meervoudige regressieanalyse uitgevoerd, waarbij *Hard Performance* eerst voorspeld is op basis van Leeftijd

en Geslacht. Dat leverde een correlatie op van 0.12 ($p = .53$). In stap 2 is daar HEXACO Consciëntieusheid aan toegevoegd. Dat leverde een multiple correlatie van 0.43 op ($\Delta R = .31$, $p = .00$). Het toevoegen van spraaktaak Consciëntieusheid als derde variabele leverde een multiple correlatie van 0.45 op ($\Delta R = .02$, $p = .16$). Voor de voorspelling van *Hard Performance* lijkt dus te kunnen worden volstaan met Consciëntieusheid zelfbeoordeling (HEXACO). De toegevoegde voorspellende waarde van de spraaktaak is zeer gering en in dit verband ook niet significant.

Tot slot is ook voor het voorspellen van *Soft Performance* een stapsgewijze meervoudige regressieanalyse uitgevoerd, waarbij *Soft Performance* eerst voorspeld is op basis van Leeftijd en Geslacht. Dat leverde een correlatie op van 0.27 ($p = .01$). In stap 2 is daar HEXACO Consciëntieusheid aan toegevoegd. Dat leverde een multiple correlatie van 0.27 op ($\Delta R = .00$, $p = .92$). Het toevoegen van spraaktaak Consciëntieusheid als derde variabele leverde een multiple correlatie van 0.31 op ($\Delta R = .04$, $p = .13$). Voor de voorspelling van *Soft Performance* lijkt dus te kunnen worden volstaan met de variabelen Leeftijd en Geslacht. Zowel Consciëntieusheid zelfbeoordeling (HEXACO) en de spraaktaak hebben nauwelijks toegevoegde voorspellende waarde en bovendien zijn beide verbanden niet significant.

Hypothese 3 kan op basis van de bovenstaande drie multiple regressie analyses worden aangenomen: De geautomatiseerde beoordelingen van Consciëntieusheid hebben geen incrementele predictieve validiteit bovenop de HEXACO zelfbeoordelingen van Consciëntieusheid.

Post-hoc analyses

De verwachting was dat het aantal woorden invloed zou hebben op de scores, omdat de kans op het aantal 'hits' groter is bij gebruik van een groter aantal woorden. Hierdoor zal de score op de persoonlijkheidsdimensie Consciëntieusheid ook beïnvloed worden. Op basis van uitgevoerde post-hoc analyses is naar voren gekomen dat voor de persoonlijkheidsdimensie Consciëntieusheid het aantal woorden niet gerelateerd is aan de scores op Consciëntieusheid ($r = -.06$). Deze uitkomst wijkt af van eerder onderzoek door Pennebaker et al. (2007), waarbij een toename van het aantal woorden gedetecteerd door het LIWC systeem een toename van de score op de persoonlijkheidsdimensie Consciëntieusheid tot gevolg had. Een mogelijke verklaring voor deze afwijkende bevinding is dat door een toename van het aantal woorden de kans op het gebruik van een negatief geladen Consciëntieusheidsadjectief ook groter is. Dit onderzoek laat, in tegenstelling tot eerder onderzoek van Pennebaker et al. (2007), zien dat het niet vanzelfsprekend is dat een toename van het aantal woorden leidt tot een hogere score op de persoonlijkheidsdimensie Consciëntieusheid.

Correlatiematrix

Tabel 1.

Gemiddelden, Standaard Deviaties, Betrouwbaarheid en Correlaties tussen de HEXACO, de Spraaktaak en Algemene werkprestatie.

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Geslacht	1.14	0.35	(-)														
2. Leeftijd	38.99	11.68	-.05	(-)													
3. HEXACO Integriteit	3.73	0.50	.09	.21*	(.69)												
4. HEXACO Emotionaliteit	2.60	0.53	.39**	.26**	.15	(.75)											
5. HEXACO Extraversie	3.74	0.43	-.11	-.21	-.01	-.31**	(.73)										
6. HEXACO Verdraagzaamheid	3.20	0.51	.15	.03	.14	.01	-.12	(.68)									
7. HEXACO Consciëntieusheid	3.71	0.49	.12	.20	.25*	.08	.05	-.04	(.76)								
8. HEXACO Openheid voor Ervaringen	3.37	0.57	-.01	-.01	-.20	.07	.05	-.02	-.08	(.73)							
9. Spraaktaak vraag 4	0.18	0.26	.02	-.02	-.00	.03	-.11	.04	-.06	-.03	(-)						
10. Totaal Spraaktaak Consciëntieusheid	0.61	0.46	.14	-.08	.24*	.15	.06	-.07	.27**	-.01	.37**	(-)					
11. Gemiddelde Spraaktaak Consciëntieusheid	0.23	0.19	.09	-.05	.23*	.04	-.04	-.10	.29**	-.03	.19	.81**	(-)				
12. Gemiddelde zonder 4 Consciëntieusheid	0.19	0.21	.07	-.01	.22*	.06	-.01	-.05	.33**	.03	-.10	.74**	.91**	(-)			
13. Mean Performance	3.29	0.33	.14	-.09	.12	-.05	.13	-.15	.37**	-.19	-.02	.22*	.24*	.26*	(.72)		
14. Hard Performance	3.35	0.43	.08	-.10	.11	-.09	.16	-.16	.39**	-.18	-.02	.21*	.21*	.24*	.97**	(.75)	
15. Soft Performance	3.13	0.28	.27	.01	.05	.17	-.10	.03	.03	-.05	-.01	.09	.15	.15	.39**	.17	(.47)

Note. $N = 90$ voor HEXACO en spraaktaak, $N = 125$ voor Algemene werkprestatie (Mean, Hard en Soft). De betrouwbaarheid van de variabelen staat op de diagonaal. M = gemiddelde. SD = standaarddeviatie. Geslacht: 1 = man, 2 = vrouw. De HEXACO variabelen hebben een minimale score van 1 en een maximum score van 5.

* $p < .05$, ** $p < .01$

Beschrijvende statistiek

Tabel 2.

Gemiddelde resultaten op alle spraaktaak persoonlijkheidsdimensies.

	M	SD
Spraaktaak 1 H	.18	.30
Spraaktaak 1 E	-.02	.27
Spraaktaak 1 X	.15	.26
Spraaktaak 1 A	.24	.31
Spraaktaak 1 C	.27	.34
Spraaktaak 1 O	.06	.28
Spraaktaak 2 H	.22	.31
Spraaktaak 2 E	-.01	.30
Spraaktaak 2 X	.23	.35
Spraaktaak 2 A	.18	.35
Spraaktaak 2 C	.27	.38
Spraaktaak 2 O	.13	.31
Spraaktaak 3 H	.12	.29
Spraaktaak 3 E	-.00	.29
Spraaktaak 3 X	.06	.31
Spraaktaak 3 A	.15	.31
Spraaktaak 3 C	.15	.35
Spraaktaak 3 O	.10	.34
Spraaktaak 4 H	.25	.28
Spraaktaak 4 E	-.02	.26
Spraaktaak 4 X	.19	.28
Spraaktaak 4 A	.29	.30
Spraaktaak 4 C	.16	.25
Spraaktaak 4 O	.09	.26
Spraaktaak 5 H	.19	.31
Spraaktaak 5 E	-.00	.29
Spraaktaak 5 X	.15	.28
Spraaktaak 5 A	.21	.32
Spraaktaak 5 C	-.01	.31
Spraaktaak 5 O	.16	.31

Note. $N = 90$. M = gemiddelde. SD = standaarddeviatie.

Regressieanalyse

Tabel 3.

Hiërarchische meervoudige regressieanalyse tussen de HEXACO, de Spraaktaak en Algemene werkprestatie.

	<i>Mean Performance</i>			<i>Hard Performance</i>			<i>Soft Performance</i>		
	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3
Geslacht	.12	.07	.07	.08	.02	.01	.21	.21	.21
Leeftijd	-.00	-.01	-.00	-.00	-.01	-.01	.00	.00	.00
Consciëntieusheid		.27**	.22**		.37**	.33**		-.01	-.04
Gemiddelde Spraaktaak zonder vraag 4			.29			.32			.23
R^2	.03	.17	.20	.01	.19	.20	.07	.07	.10
ΔR^2	.03	.15	.03	.01	.17	.02	.07	.00	.03
ΔF	1.10	15.12	2.85	0.60	17.93	1.99	3.34	0.01	2.32
Sig. ΔF	.34	.00	.10	.55	.00	.16	.04	.92	.13

Note. $N = 90$ voor HEXACO en spraaktaak, $N = 90$ voor Algemene werkprestatie (Mean, Hard en Soft). Gestandaardiseerde regressiecoëfficiënten (bèta) worden getoond in de eerste vier rijen. ΔR^2 kan inconsistent blijken als gevolg van afronding. Stap 1: Leeftijd en Geslacht, Stap 2: HEXACO-PI Consciëntieusheid, Stap 3: Spraaktaak Consciëntieusheid.

* $p < .05$, ** $p < .01$

Discussie

Dit onderzoek heeft zich gericht op de vraag in hoeverre het geautomatiseerd beoordelen van (gesproken) tekst menselijke zelfbeoordelingen kan vervangen of aanvullen om werkprestatie te voorspellen. Dit is op twee manieren onderzocht. Als eerste is de constructvaliditeit van de een geautomatiseerde beoordeling van persoonlijkheid aangetoond. Ten tweede is de predictieve validiteit van deze geautomatiseerde beoordeling aangetoond.

De onderzoeksresultaten ondersteunen allereerst de relatie tussen de geautomatiseerde beoordelingen en de zelfbeoordelingen. Dit resultaat komt overeen met de verwachting die is geformuleerd in hypothese 1. Ook hebben de resultaten van dit onderzoek, overeenkomstig met de verwachting geformuleerd in hypothese 2a, aangetoond dat Consciëntieusheid zelfbeoordeling de mate van werkprestatie kan voorspellen. Daarnaast is er een significant verband gevonden tussen geautomatiseerde beoordelingen van tekst over Consciëntieusheid en Algemene werkprestatie. Echter, spraaktaal Consciëntieusheid had geen incrementele predictieve validiteit ten opzichte van de HEXACO Consciëntieusheid. Overigens heeft het aantal woorden dat een respondent gebruikt tijdens de spraaktaken geen invloed op de score op de persoonlijkheidsdimensies.

Het aannemen van hypothese 1 betekent dat, ondanks de eerder beschreven lage alfa betrouwbaarheden, de spraaktaal een mate van constructvaliditeit heeft. De persoonlijkheidsdimensie Consciëntieusheid wordt immers ook voor een deel door de spraaktaal gemeten, en dat was een onderdeel van dit onderzoek. Vraag 4 moest echter wel uit de spraaktaal verwijderd worden om dit resultaat te bereiken. Bovendien is het verband tussen beide variabelen niet heel groot, want het percentage verklaarde variantie is 10,9% ($r^2 = 0.109$).

Dit betekent dat 10,9% van zelf beoordeeld Consciëntieusheid door de spraaktaak Consciëntieusheid verklaard wordt. Deze werkelijke samenhang is niet zo sterk als verwacht. Gebaseerd op eerdere LIWC onderzoeken, waarin de persoonlijkheid van de deelnemers geautomatiseerd beoordeeld wordt (Mairesse et al., 2007), werd verwacht dat de samenhang tussen de $r = .44$ en 0.54 zou zijn. Het speciaal ontwikkelde algoritme is namelijk geavanceerder dan het LIWC dat in voorgaande onderzoeken werd gebruikt, waardoor er een sterke samenhang verwacht werd. Uit de resultaten van dit onderzoek komt een Pearson's correlatiecoëfficiënt van 0.33 naar voren. Een verklaring hiervoor zou kunnen zijn dat de input voor het speciaal ontwikkelde algoritme nog niet optimaal is. In de instructie van de spraaktaak staat vermeld dat respondenten zo veel mogelijk bijvoeglijk naamwoorden moeten gebruiken om zichzelf te omschrijven. Het algoritme herkent namelijk alleen vooraf ingevoerde bijvoeglijk naamwoorden die samenhangen met de verschillende persoonlijkheidsdimensies. Een veel gehoorde opmerking van de respondenten was dat zij het lastig vonden om zichzelf te omschrijven aan de hand van bijvoeglijk naamwoorden. Dit kan leiden tot een beperkte meting, wanneer kandidaten geen bijvoeglijknaamwoorden verzinnen. Het algoritme is namelijk niet in staat om - buiten de gebruikte bijvoeglijk naamwoorden om - persoonlijkheidsdimensies te detecteren, terwijl dit wel uit de woorden van de respondenten herleid zou kunnen worden. Daarnaast is er gekozen voor minder letterlijke vragen, omdat dit over het algemeen tot meer persoonlijkheid gerelateerde antwoorden leidt (Van Iddekinge et al., 2005). Op basis van de feedback van de respondenten komt naar voren dat veel respondenten de vragen niet duidelijk genoeg vonden en dat zij niet precies wisten hoe hun antwoord eruit moest zien. Respondenten kwamen vaak met de suggestie om de vragen te stellen aan de hand van een bepaalde case, zodat zij zich makkelijker in de situatie kunnen verplaatsen. Hiervoor is in dit onderzoek juist niet gekozen, omdat dit soort

vragen minder bijvoeglijknaamwoorden oproepen. Daarbij komt het feit dat er vooral mannen met een technische functie binnen een ingenieursbureau hebben deelgenomen aan dit onderzoek. Tijdens het uitschrijven van de spraaktaken viel het namelijk op dat de mannelijke respondenten het over het algemeen lastig vonden om zichzelf te omschrijven in abstracte situaties. Er werd vooral procesmatig naar het eigen functioneren gekeken, maar er werd nauwelijks omschreven naar welk gedrag de respondenten neigen in bepaalde situaties. Tot slot kwam uit de feedback naar voren dat respondenten het vreemd vonden om in een lege ruimte tegen een computer te praten. Veel respondenten gaven aan 'dicht te klappen' en gaven de voorkeur aan een face-to-face gesprek om iets over zichzelf te vertellen.

Het aannemen van hypothese 2a komt overeen met de verwachting, gebaseerd op eerder wetenschappelijk onderzoek dat Consciëntieusheid een goede voorspeller is van succesvolle prestaties op de werkvloer bij werknemers (bv. Barrick & Mount, 2006; Salgado, 1997; Higgins et al, 2007). Dit bevestigt het nut van het meten van persoonlijkheid om in te kunnen schatten of een kandidaat geschikt is voor een bepaalde functie. Het significant verband tussen enerzijds Consciëntieusheid zelfbeoordeling (HEXACO) en anderzijds Algemene werkprestatie ($r = .37$) en Hard Performance ($r = .39$) is op basis van de algemene vuistregels is zwak. Echter, voor assessment psychologie is dit significante verband helemaal niet zwak. Op basis van meta-analyse correlaties uitgevoerd door Schmidt en Hunter (1998) blijkt dat persoonlijkheidsvragenlijsten een voorspellende validiteit hebben van .31. In verhouding met deze correlatie is de correlatie in dit onderzoek hoog.

Op basis van de hiërarchische meervoudige regressieanalyse, toegepast op drie verschillende criteria (*Mean Performance*, *Hard Performance* en *Soft Performance*), is een nieuw inzicht naar voren gekomen. Er blijkt een significant verband te zijn tussen

Consciëntieusheid zelfbeoordeling enerzijds en zowel *Mean Performance* ($r = .42, p < .01$) als *Hard Performance* ($r = .43, p < .01$) anderzijds. Echter, er is geen significant verband gevonden tussen Consciëntieusheid zelfbeoordeling en *Soft Performance*. De onderdelen *Integrity* en *Health, Safety & Environment* zouden dus op een andere manier gemeten moeten worden om in te kunnen schatten hoe een kandidaat op dit onderdeel scoort.

Hypothese 2b is eveneens aangenomen, er is een verband tussen de geautomatiseerde beoordelingen en Algemene werkprestatie. Echter, Consciëntieusheid zelfbeoordelingen hebben een grotere voorspellende waarde voor werkprestatie dan Geautomatiseerde Consciëntieusheidsbeoordelingen. Geautomatiseerde beoordelingen van Consciëntieusheid hebben geen toegevoegde voorspellende waarde bovenop de Consciëntieusheid zelfbeoordelingen. Het was al de verwachting dat Consciëntieusheid zelfbeoordeling een betere voorspellende waarde heeft, omdat de validiteit en de betrouwbaarheid van zelfbeoordelingsvragenlijsten al meerdere malen wetenschappelijk zijn aangetoond (De Vries, 2006).

Theoretische en praktische implicaties

Dit onderzoek heeft een bijdrage geleverd aan de literatuur op het gebied van het detecteren van persoonlijkheid aan de hand van (gesproken) tekst. Niet eerder werd daarbij een relatie gemaakt tussen het detecteren van persoonlijkheid enerzijds en het voorspellen van werkprestatie door middel van de persoonlijkheidsdimensie Consciëntieusheid anderzijds. Uit meerdere onderzoeken blijkt dat persoonlijkheidsdimensies gemeten kunnen worden op basis van een stuk tekst en dat Consciëntieusheid de beste voorspeller is van werkprestatie. Dit onderzoek brengt nieuwe methoden en vastgestelde assessmentprincipes bij elkaar. Tevens draagt dit onderzoek bij

aan het verzoek van Hirsch en Peterson (2009) om een geavanceerd tekstanalyse softwaresysteem te ontwikkelen dat tegemoet komt aan de beperkingen van het veel gebruikte LIWC systeem. Deze informatie zou kunnen leiden tot voldoende handvatten voor andere wetenschappers om meer onderzoek te doen op dit gebied.

Het aannemen van hypothese 2a bevestigt het nut van het meten van persoonlijkheid om in te kunnen schatten of een kandidaat geschikt is voor een bepaalde functie. Op basis van data verzameld bij een groot ingenieursbureau, laat dit onderzoek zien dat er een significant verband is tussen de persoonlijkheidsdimensie Consciëntieusheid en Algemene werkprestatie. Organisaties in het algemeen en Human Resources afdelingen in het bijzonder zouden door middel van dit onderzoek het belang van het meten van persoonlijkheid om werkprestatie te kunnen voorspellen kunnen onderstrepen. Alhoewel dit in eerdere onderzoeken ook al bewezen is, brengt dit onderzoek toch nieuwe kennis met zich mee. Wat interessant is, is dat men met het meten van de persoonlijkheidsdimensie Consciëntieusheid vooral *Hard Performance* kan voorspellen.

Het aannemen van hypothese 3 betekent niet dat het gebruik van de spraaktaak nutteloos is binnen organisaties. Alhoewel wetenschappelijk niet bewezen is dat de spraaktaak een incrementele validiteit heeft, is het wel mogelijk dat deelnemers aan de spraaktaak deze test als meer relevant ervaren voor het bepalen van het niveau van werkrelevante kennis, vaardigheden en kwaliteiten in vergelijking met andere selectie instrumenten. Dit wordt ook wel indrukvaliditeit genoemd (Cook, 2004). Een hoge indrukvaliditeit zorgt ervoor dat de vragen van het selectie instrument serieuzer worden genomen, waardoor er meer aandacht aan deze vragen besteed wordt en de vragen eerlijker worden ingevuld in vergelijking met een selectie instrument met een lagere indrukvaliditeit. Alhoewel wetenschappelijk gezien de

indruksvaliditeit nauwelijks of geen waarde heeft, worden er in de praktijk echter regelmatig selectie instrumenten gebruikt omdat die een hoge mate van indruksvaliditeit hebben. In dit geval zal het gebruik van spraaktaak Consciëntieusheid alsnog een toegevoegde waarde kunnen hebben ten opzichte van HEXACO Consciëntieusheid.

Beperkingen en suggesties voor vervolgonderzoek

Zoals de meeste wetenschappelijke onderzoeken, heeft ook dit onderzoek beperkingen. Allereerst is er een methodologische beperking bij de spraaktaak. Het was een uitdaging om vragen te formuleren om de persoonlijkheidsdimensie Consciëntieusheid te meten en tegelijk respondenten zover te krijgen om zichzelf te omschrijven aan de hand van bijvoeglijk naamwoorden. Er is hier mogelijk een *common source bias* (Podsakoff et al., 2003) opgetreden, waarbij de vertekeningen in de resultaten veroorzaakt zijn door de studie zelf in plaats van de werkelijke situatie. De *bias* kan ontstaan door de manier waarop de vragen zijn samengesteld, de manier waarop de vragen zijn gesteld en het soort respondenten waaraan de vragen zijn gesteld. Een deel van de vragen zijn namelijk samengesteld op basis van de *Core Skills* van Fluor. Deze *Core Skills Accountability* en *Drive for Results* zijn conceptueel sterk gerelateerd aan Consciëntieusheid. De prestatiebeoordelingen lijken dus al gedeeltelijk Consciëntieusheid gerelateerd te zijn. Daarnaast moesten de vragen zó gesteld worden, dat de respondenten gestimuleerd werden om zo veel mogelijk bijvoeglijk naamwoorden te gebruiken. Dit resulteerde in moeilijke te begrijpen vragen voor sommige respondenten.

Daarnaast heeft het onderzoek de vorm van een cross-sectioneel onderzoek. Alle variabelen worden gelijktijdig onderzocht. Dit brengt een beperking met zich mee, omdat dit onderzoek gedrag in de toekomst probeert te voorspellen op basis van informatie die nu

beschikbaar is. Een longitudinaal onderzoek zou beter op zijn plek zijn, alhoewel dit ook weer nadelen met zich meebrengt. Het is immers tijdrovend en de kans bestaat dat de samenstelling van de respondenten op den duur verandert, waardoor sommige uitkomsten onbruikbaar zijn.

Tot slot moet er een kanttekening geplaatst worden bij de generaliseerbaarheid van de resultaten. De respondenten die hebben deelgenomen aan dit onderzoek zijn allen werkzaam bij een groot ingenieursbureau. Het aantal mannen ten opzichte van het aantal vrouwen is hier niet normaal verdeeld. Aan dit onderzoek hebben 109 mannen en 16 vrouwen deelgenomen. Deze steekproef is niet representatief voor een grote bevolkingsgroep.

Vervolgonderzoek zou zich vooral moeten richten op het door ontwikkelen van een tekstanalyse programma dat tegemoet komt aan de beperkingen van het LIWC systeem en het algoritme Sentimentics. Omdat bij beiden tekstanalyse programma's de detectie van persoonlijkheid beperkt is tot het herkennen van woorden die vooraf aan het programma worden meegegeven, wordt er nog veel waardevolle informatie 'gemist' en niet meegenomen in de resultaten van de onderzoeken. Daarnaast zou vervolgonderzoek zich kunnen richten op het betrouwbaar maken van de metingen van de spraaktaak. Een mogelijk oorzaak van de lage alfa betrouwbaarheid is de geringe bruikbare data ($N=90$) die beschikbaar was voor dit onderzoek. Vervolgonderzoek zou zich kunnen richten op het vergroten van de steekproef, waardoor de kans groter is dat er wel goede alfa betrouwbaarheden worden gevonden.

Ook kan vervolgonderzoek zich richten op het beoordelen van de spraaktaak. De grote uitdaging voor deze spraaktaak was het bedenken van de vragen waarbij de respondenten enerzijds gestimuleerd moesten worden om over hun persoonlijkheid te gaan praten en anderzijds gebruik moesten maken van bijvoeglijk naamwoorden, zodat het algoritme een score

kan berekenen. In de instructie werd aangegeven dat het gebruik van bijvoeglijk naamwoorden aangemoedigd werd en werden er een aantal voorbeelden gegeven. Voor dit onderzoek werden er algemene vragen in een werkcontext gesteld, omdat uit de literatuur blijkt dat de persoonlijkheid van mensen dan het best naar voren komt. Veel respondenten gaven aan dat ze het lastig vonden om te antwoorden op algemene vragen en wilden liever een concrete situatie waarin ze zichzelf moesten beschrijven. Een mogelijke manier van vragen zou kunnen zijn dat de vragen van de spraaktaak worden aangepast aan de situatie van de individuele kandidaat, waardoor de kandidaat zich beter kan verplaatsen in die situatie en dus makkelijker over zijn of haar gedrag in die situatie kan praten.

Een andere suggestie voor vervolgonderzoek is het ontwikkelen van een betrouwbare *speech-to-text* applicatie dat de spraak tijdens een spraaktaak direct omzet in tekst, waarna de tekst door het algoritme geanalyseerd wordt. Dit kan tot een verbetering van de onderzoek tijd leiden, omdat voor dit onderzoek alle 103, waarvan 90 bruikbare, spraaktaken woord voor woord uitgeschreven moesten worden. Voor dit onderzoek is er in eerste instantie gebruik gemaakt van een *speech-to-text* applicatie, maar de uitkomsten van de applicatie waren niet betrouwbaar genoeg om mee te nemen in de analyses.

Tot slot zou vervolgonderzoek zich kunnen richten op het onderzoeken of zelfbeoordelaars *biases* echt niet terugkomen bij geautomatiseerde beoordelingen. Alhoewel er geen zelfbeoordelvragenlijsten gemaakt worden en de kans op *response bias* daarom klein is, hebben respondenten wel de kans om een positief zelfbeeld van zichzelf te creëren om de resultaten te beïnvloeden. De meest voorkomende invulneigingen tijdens het maken van een (zelfbeoordelings)vragenlijsten (He & Van de Vijver, 2013) kunnen ook van toepassing zijn op het beantwoorden van vragen tijdens een spraaktaak. Het gaat hier om *acquiescence* (de neiging

om het eens te zijn met items, ongeacht de inhoud daarvan), *extremity* (de neiging om de uitersten van de schalen veel te gebruiken) en *midpoint responding* (de neiging om het midden van een schaal veel te gebruiken).

Conclusie

Dit onderzoek heeft als doel een antwoord te vinden op de vraag in hoeverre het geautomatiseerd beoordelen van (gesproken) tekst bestaande beoordelingen kan vervangen of aanvullen om werkprestatie te voorspellen. In dit onderzoek is aangetoond dat de resultaten van de speciaal ontwikkelde spraaktaak correleren met de resultaten van de wetenschappelijk bewezen HEXACO-PI vragenlijst en dat zowel Consciëntieusheid zelfbeoordeling als Consciëntieusheid spraaktaak een predictieve validiteit heeft in relatie met Algemene werkprestatie. Echter, dit onderzoek heeft tevens aangetoond dat het geautomatiseerd beoordelen van (gesproken) tekst bestaande beoordelingen niet kan vervangen of aanvullen om werkprestatie te voorspellen.

Dit onderzoek is een goede stap op weg naar de mogelijkheid om met zo min mogelijk *biases* beoordelingen te kunnen geven. Verder onderzoek zal nodig zijn om het algoritme geavanceerder te maken, zodat er een grotere kans is dat de antwoorden van de respondenten op de spraaktaak herkend worden.

Referenties

- Ashton, M. C. (1998). Personality and job performance: The importance of narrow traits. *Journal of Organizational Behavior, 19*, 289-303.
- Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*, 150-166.
- Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340-345.
- Baarda, B. (2014). *Dit is onderzoek! Handleiding voor het kwantitatief en kwalitatief onderzoek*. Groningen: Noordhoff Uitgevers
- Barrick, M. R., & Mount, M.K. (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis, *Personnel Psychology, 44*, 1-26.
- Borman, W. C, Ilgen, D. R., & Klimoski, R. J. (2003). Industrial and Organizational Psychology. In I. B. Weiner (Ed.), *Handbook of Psychology* (pp. 77-105). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Burisch, M. (1997). Test length and validity revisited. *European Journal of Personality, 11*, 303-315.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England: Cambridge University Press.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, New Jersey: Erlbaum.
- Cook, M. (2004). *Personnel selection: Adding value through people* (4th ed.). Chichester: John Wiley & Sons Ltd.

- Costa, P. T., & McCrae, R.R. (1988). Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology*, *54*, 853-863.
- Credé, M., Bashshur, M., & Niehorster, S. (2010). Reference group effects in the measurement of personality and attitudes. *Journal of Personality Assessment*, *92*, 390-399.
- Dewitt, S., & Schouwenburg, H. C. (2002). Procrastination, temptations, and incentives: The struggle between the present and the future in procrastinators and the punctual. *European Journal of Personality*, *16*, 469-489.
- Dodorico-McDonald, J. (2008). Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire*, *1*, 1-19.
- Eder, R. W., & Harris, M. M. (Eds.) (1999). *The employment Interview handbook*. Newbury Park, CA: SAGE Publications, Inc.
- Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance-report, and behavior. *Journal of Personality and Social Psychology*, *94*, 334-346.
- Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*, 1216-1229.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits: Authors' reactions to the six comments. *American Psychologist*, *48*, 1303-1304.
- Haselton, M. G., Nettle, D., & Andrews, P. W. (2005). The evolution of cognitive bias. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 724–746). Hoboken, New Jersey: John Wiley & Sons Inc.

- He, J., & Van de Vijver, F. J. (2013). A general response style factor: Evidence from a multi-ethnic study in the Netherlands. *Personality and Individual Differences, 55*, 794-800.
- Higgins, D. M., Peterson, J.B., Lee, A., & Pihl, R.O. (2007). Prefrontal cognitive ability, intelligence, Big Five personality and the prediction of advanced academic and workplace performance. *Journal of Personality and Social Psychology, 93*, 298-319.
- Hirsch, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality, 43*, 524-527.
- Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and Answers. *American Psychologist, 51*, 469-477.
- Hurtz, G. M., & Donovan, J. J. (2000) Personality and Job Performance: The Big Five Revisited. *Journal of Applied Psychology, 85*, 869-879.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996). Bias in Judgment: Comparing Individuals and Groups. *Psychological Review, 103*, 687-719.
- Kleinmann, M., Ingold, P. V., Lievens, F., Jansen, A., Melchers, K. G., & König, C. J. (2011). A different look at why selection procedures work: The role of candidates' ability to identify criteria. *Organizational Psychology Review, 1*, 128-146.
- Lewis, A. (2008). *The Cambridge Handbook of Psychology and Economic Behaviour*. Cambridge University Press.
- Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research, 30*, 457-500.

- McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as human universal. *American Psychologist*, *52*, 509-516.
- Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*, 862-877.
- Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, *84*, 857-870.
- Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-factor model of personality and Performance in jobs involving interpersonal interactions. *Human Performance*, *11*, 145-165.
- Murphy, K. R. (1996). Individual differences and behavior in organizations: Much more than *g*. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 3–30). San Francisco: Jossey-Bass.
- Oberlander, J., & Gill, A. J. (2006). Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discourse processes*, *42*, 239-270.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of Social Desirability in Personality Testing for Personnel Selection: The Red Herring. *Journal of Applied Psychology*, *81*, 660-679.
- Paulhus, D. L. (2002). Socially Desirable Responding: The Evolution of a Construct. In: Braun, H. I., Jackson, D. N., & Wiley, D. E. (Eds.), *The Role of Constructs in Psychological and Educational Measurement* (pp. 49-69). Mahwah NJ: Erlbaum.

- Paunonen, S. V., & Jackson, D. N. (1985). The validity of formal and informal personality assessments. *Journal of Research in Personality, 19*, 331-342.
- Pennebaker, J. W. (1993). Putting stress into words: Health, linguistic, and therapeutic implications. *Behaviour Research and Therapy, 31*, 539-548.
- Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007). *Linguistic Inquiry and Word Count (LIWC2007): A computer-based text analysis program [Computer software]*. Austin, TX: LIWC.net.
- Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition and Emotion, 10*, 601-626.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*, 1296-1312.
- Pennebaker, J. W., Mayne, T., & Francis, M. (1997). Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology, 72*, 863-871.
- Pfister, H. R., & Böhm, G. (2008). The multiplicity of emotions: A framework of emotional functions in decision making. *Judgment and Decision Making, 3*, 5-17.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903.
- Risavy, S. D., & Hausdorf, P. A. (2011). Personality Testing in Personnel Selection: Adverse impact and differential hiring rates. *International Journal of Selection and Assessment, 19*, 18-30.

- Roberts, B. W., Jackson, J. J., Fayard, J.V., Edmonds, G., & Meints, J (2009). Chapter 25. Conscientiousness. In M. R. Leary, & R. H. Hoyle (pp. 257–273). *Handbook of Individual Differences in Social Behavior*. New York/London: The Guildford Press.
- Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support?. *Human Resource Management Review*, 16, 155.
- Ruz, M., Moser, A., & Webster, K. (2011). Social Expectations Bias Decision-Making in Uncertain Inter-Personal Situations. *PLoS ONE*, 6 (2): e15762.
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology*, 82, 30-43.
- Schmidt, F. L., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and Theoretical Implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmidt, F. L., & Hunter, J. (2004). General Mental Ability in the World of Work: Occupational Attainment and Job Performance. *Journal of Personality and Social Psychology*, 86, 162-173.
- Segrest-Purkiss, S. L., Perrewé, P. L., Gillespie, T. L., Mayes, B. T., & Ferris, G. R. (2006). Implicit sources of bias in employment interview judgments and decisions. *Organizational Behavior and Human Decision Processes*, 101, 152-167.
- Sentimentics (2014). Verkregen op 21 augustus, 2014 via website: <http://sentimentics.com/>.
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69, 99-118.
- Spearman, C. (1927). *The abilities of man*. New York: MacMillan.

- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis. *Journal of Language and Social Psychology, 29*, 24-54.
- Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology, 90*, 536-552.
- Verschueren, K., & Koomen, H. (2007). *Handboek Diagnostiek in de Leerlingenbegeleiding*. Antwerpen: Garant.
- Vries, R. E. de (2013). The 24-item Brief HEXACO Inventory (BHI). *Journal of Research in Personality, 47*, 871-880.
- Vries, R. E. de, Ashton, M. C., & Lee, K. (2009). De zes belangrijkste persoonlijkheidsdimensies en de HEXACO Persoonlijkheidsvragenlijst. *Gedrag en Organisatie, 22*, 232-274.
- Wang, X. T., Simons, F., & Brédart, S. (2001). Social cues and verbal framing in risky choice. *Journal of Behavioral Decision Making, 14*, 1-15.
- Yarkoni, T. (2010). Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality, 44*, 363-373.
- Zijlstra, H., Middendorp, H. van, Meerveld, T. van, & Geenen, R. (2005) Validiteit van de Nederlandse versie van de Linguistic Inquiry and Word Count (liwc). *Netherland Journal of Psychology, 60*, 50-58.

Bijlage 1. Factoranalyses

Tabel 4.

Resultaten van de factoranalyse Spraaktaken.

	1	2	3	4	5	6
Spraaktaak 1, score Integriteit						.49
Spraaktaak 2, score Integriteit					.77	
Spraaktaak 3, score Integriteit			.92			
Spraaktaak 4, score Integriteit	.93					
Spraaktaak 5, score Integriteit		.94				
Spraaktaak 1, score Emotionaliteit				.85		
Spraaktaak 2, score Emotionaliteit						.31
Spraaktaak 3, score Emotionaliteit				.35		
Spraaktaak 4, score Emotionaliteit						
Spraaktaak 5, score Emotionaliteit					.34	
Spraaktaak 1, score Extraversie						.79
Spraaktaak 2, score Extraversie					.74	
Spraaktaak 3, score Extraversie			.81			
Spraaktaak 4, score Extraversie	.82					
Spraaktaak 5, score Extraversie		.92				
Spraaktaak 1, score Verdraagzaamheid						.85
Spraaktaak 2, score Verdraagzaamheid					.63	
Spraaktaak 3, score Verdraagzaamheid			.87			
Spraaktaak 4, score Verdraagzaamheid	.90					
Spraaktaak 5, score Verdraagzaamheid		.89				
Spraaktaak 1, score Consciëntieusheid						.79
Spraaktaak 2, score Consciëntieusheid					.61	.36
Spraaktaak 3, score Consciëntieusheid			.52			
Spraaktaak 4, score Consciëntieusheid	.69					
Spraaktaak 5, score Consciëntieusheid				.49		.39
Spraaktaak 1, score Openheid voor Ervaringen						
Spraaktaak 2, score Openheid voor Ervaringen					.64	
Spraaktaak 3, score Openheid voor Ervaringen			.61			
Spraaktaak 4, score Openheid voor Ervaringen	.69					
Spraaktaak 5, score Openheid voor Ervaringen		.91				

Note. Alleen factorladingen groter dan .30 zijn weergegeven. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

Tabel 5.

Resultaten van de factoranalyse Algemene werkprestatie.

	1	2
Excellence	.86	
Health, Safety & Environmental		.81
Integrity		.82
Teamwork	.64	
Knowledge of Work	.69	
Quality of work	.52	
Quantity of Work	.80	

Note. Alleen factorladingen groter dan .30 zijn weergegeven.
Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Bijlage 2. HEXACO-PI-60**Inleiding****Welkom!**

Hartelijk dank voor het meewerken aan dit onderzoek naar persoonlijkheid en gesproken tekst. Het doel van dit onderzoek is om meer inzicht te krijgen in een nieuwe assessment technologie.

Wanneer u deelneemt aan dit onderzoek dan is het mogelijk om een rapportje te ontvangen van uw testresultaten op de persoonlijkheidsvragenlijst. U kunt op het einde aangeven of u dit graag zou willen.

Deelname duurt ongeveer 10 minuten. Indien u het onderzoek niet in één keer kunt afronden, dan kunt u later weer doorgaan waar u gebleven was.

Alvast hartelijk dank voor uw medewerking!

Jasper Stoker
jasper.stoker@fluor.com

Bij dit onderzoek wordt volgens de Beroepscode van het NIP (www.psynip.nl) gewerkt. Alle informatie is volledig vertrouwelijk, zal alleen voor onderzoeksdoeleinden worden gebruikt en nooit met derden worden gedeeld, dus ook niet met uw werkgever.

Persoonsgegevens

Uw naam wordt alleen gebruikt om de verschillende vragenlijsten aan elkaar te koppelen, hierna zal uw naam worden verwijderd

Voornaam
Tussenvoegsel
Achternaam
Naam lead of directe leidinggevende

Geslacht

- Man
 Vrouw

Leeftijd (in jaren)

.....

Geboorteland

.....

Geboorteland moeder

.....

Geboorteland vader

.....

Hoogste afgeronde of huidige opleiding

- Basisonderwijs (lagere school)
- Lager beroepsonderwijs (vmbo, lbo of vergelijkbaar)
- Middelbaar beroepsonderwijs (mbo/mulo/lbgo/lto of vergelijkbaar)
- Hoger algemeen voortgezet onderwijs (havo)
- Voortgezet wetenschappelijk onderwijs (vwo)
- Hoger beroepsonderwijs (hbo)
- Wetenschappelijk onderwijs (universitair)
- Anders, namelijk

Hoeveel jaar bent u op dit moment actief binnen uw vakgebied?

.....

Op welke afdeling werkt u en wat is uw functie?

.....

Persoonlijkheidsvragenlijst over uzelf

Instructie:

Op de volgende pagina's vindt u uitspraken over uzelf. U wordt verzocht de uitspraken te lezen en aan te geven in hoeverre u het met deze uitspraken eens dan wel oneens bent.

U kunt bij het beantwoorden van de vragen gebruik maken van de volgende antwoordcategorieën:

- 1=Helemaal mee oneens
- 2=Mee oneens
- 3=Neutraal (Noch mee eens, noch mee oneens)
- 4=Mee eens
- 5=Helemaal mee eens

Klik het bolletje aan bij het antwoord dat u het beste bij uzelf vindt passen.

Nr.	Stelling	1	2	3	4	5
1	Ik zou me vervelen bij een bezoek aan een kunstgalerie.					
2	Ik maak vooraf plannen en regel alvast zaken om te vermijden dat ik op het laatste moment nog dingen moet doen.					
3	Ik houd zelden een wrok tegen iemand, zelfs niet als ik erg slecht behandeld ben.					
4	Alles bij elkaar heb ik wel een tevreden gevoel over mijzelf.					
5	Ik zou bang worden als ik in slecht weer zou moeten reizen.					
6	Ik zou niet vleien om op het werk opslag of promotie te krijgen, zelfs al zou het succes hebben.					
7	Ik kom graag meer te weten over de geschiedenis en politiek van andere landen.					
8	Ik span me vaak tot het uiterste in als ik een doel tracht te bereiken.					
9	Mensen vertellen me soms dat ik te kritisch op anderen ben.					
10	Ik geef zelden mijn mening in groepsbijeenkomsten.					
11	Ik maak me soms zorgen over onbenulligheden.					
12	Als ik niet gepakt zou worden, dan zou ik er geen probleem mee hebben om een miljoen					

	Euro te stelen.					
13	Ik zou graag iets kunstzinnigs doen, zoals een boek schrijven, een lied componeren of een schilderij maken.					
14	Als ik aan iets werk, besteed ik weinig aandacht aan kleine details.					
15	Mensen vertellen me soms dat ik te koppig ben.					
16	Ik heb liever een baan waarin men veel met andere mensen omgaat dan één waarin men alleen dient te werken.					
17	Na een pijnlijke ervaring heb ik iemand nodig om me te troosten.					
18	Veel geld bezitten vind ik onbelangrijk.					
19	Ik vind het tijdverlies om aandacht te besteden aan radicale ideeën.					
20	Ik neem beslissingen op basis van 'hier-en-nu' gevoelens in plaats van zorgvuldig beraad.					
21	Mensen vinden me een heethoofd.					
22	De meeste dagen voel ik me blij en optimistisch.					
23	Ik voel tranen opkomen als ik anderen zie huilen.					
24	Ik vind dat ik meer recht op respect heb dan de gemiddelde persoon.					
25	Als ik de gelegenheid had, zou ik graag een klassiek concert bijwonen.					
26	Ik haal me soms problemen op de hals omdat ik slordig ben.					
27	Mijn houding ten aanzien van mensen die mij slecht behandeld hebben is "vergeven en					

	vergeten".					
28	Ik heb het gevoel dat ik een impopulair persoon ben.					
29	Als het gaat om fysiek gevaar, ben ik een angstaas.					
30	Als ik iets van iemand wil, lach ik om diens slechtste grappen.					
31	Ik heb nooit met veel plezier in een encyclopedie gekeken.					
32	Ik verricht zo min mogelijk werk, maar net genoeg om rond te komen.					
33	Ik heb de neiging andere mensen mild te beoordelen.					
34	Als ik anderen ontmoet, ben ik meestal diegene die het contact op gang brengt.					
35	Ik maak me veel minder zorgen dan de meeste mensen.					
36	Ik zou nooit ingaan op een poging tot omkoping, zelfs niet als het om een erg hoog bedrag ging.					
37	Mensen vertellen me vaak dat ik een levendige verbeelding heb.					
38	Ik probeer altijd zo nauwkeurig mogelijk te werken, zelfs al kost het me extra tijd.					
39	Ik ben gewoonlijk vrij flexibel in mijn opvattingen als mensen het met mij oneens zijn.					
40	Het eerste dat ik altijd doe als ik ergens nieuw ben, is vrienden maken.					
41	Moeilijke situaties kan ik aan zonder emotionele steun van anderen nodig te hebben.					
42	Ik zou veel plezier beleven aan het bezit van dure luxe goederen.					
43	Ik houd wel van mensen met onconventionele ideeën.					
44	Ik maak veel fouten omdat ik niet nadenk voordat ik iets doe.					

45	De meeste mensen hebben de neiging sneller boos te worden dan ik.					
46	De meeste mensen zijn levenslustiger en dynamischer dan ik over het algemeen ben.					
47	Ik raak erg geëmotioneerd als iemand die me na staat voor een lange tijd weg gaat.					
48	Ik wil dat mensen weten hoe belangrijk ik ben.					
49	Ik beschouw mezelf niet als een artistiek of creatief type.					
50	Mensen noemen me vaak een perfectionist.					
51	Zelfs als mensen veel fouten maken, zeg ik zelden iets negatiefs.					
52	Soms heb ik het gevoel dat ik een waardeloos persoon ben.					
53	Zelfs in crisissituaties blijf ik rustig.					
54	Ik zou niet net doen alsof ik iemand mag om te zorgen dat die persoon mij een dienst bewijst.					
55	Ik vind het saai om over filosofie te discussiëren.					
56	Ik doe liever dingen spontaan dan vast te houden aan een plan.					
57	Als mensen mij vertellen dat ik het mis heb, is mijn eerste reactie dit aan te vechten.					
58	Als ik met andere mensen samen ben, ben ik vaak de woordvoerder van de groep.					
59	Ik raak niet snel geëmotioneerd, zelfs niet in situaties waarin anderen erg sentimenteel worden.					
60	Ik zou in de verleiding komen om vals geld te gebruiken als ik er zeker van was dat ik er mee weg zou komen.					

Bijlage 3. Spraaktaak

Instructie:

Bedankt voor uw deelname aan dit onderzoek. Het doel van het onderzoek is om meer inzicht te krijgen in een nieuwe assessment technologie. Het onderzoek zal ongeveer 10 minuten in beslag nemen. Deze spraaktaak bestaat uit vijf open vragen met een beschikbare tijd per vraag van twee minuten. De bedoeling is om de getoonde vragen zo uitgebreid mogelijk te beantwoorden. Onderin verschijnt per vraag een tijdbalk en ik zou u willen vragen om, indien mogelijk, de volledig beschikbare tijd per vraag te gebruiken. Deze taak is geen 'toets'. Er zijn derhalve geen goede of foute antwoorden. Om uw input zo goed mogelijk te kunnen analyseren, wordt uw antwoord op de vragen opgenomen. Uw opname kan opgeslagen worden onder een tweecijferige code die vooraf aan de spraaktaak door de onderzoeker wordt medegedeeld. De onderzoeksgegevens worden daarna volledig anoniem geanalyseerd. Fluor krijgt geen inzicht in uw persoonlijke antwoorden en/of resultaten.

Vanaf de volgende pagina verschijnen de vragen. U krijgt steeds een situatie voorgelegd, waarbij u zich moet omschrijven als persoon. De bedoeling is om zoveel mogelijk *bijvoeglijk naamwoorden* te gebruiken die op u van toepassing zijn en uit te leggen waarom. Voorbeelden van mogelijke woorden zijn *zorgvuldig, ordelijk, impulsief, flexibel, laks, gemakzuchtig, ambitieus, nonchalant, verantwoordelijk, onverschillig*, enzovoorts. Ik moedig u aan om ook eigen bijvoeglijk naamwoorden te gebruiken om uzelf te omschrijven. Lees de vragen rustig door en druk op spatie om te beginnen met antwoorden op de vraag. Zodra de tijdbalk verdwenen is, kunt u weer op spatie drukken om naar de volgende vraag te gaan.

Veel plezier!

Vraag 1:

U krijgt een aantal taken opgelegd voor een opdracht of project. Hoe zorgt u ervoor dat de opgelegde taken op een efficiënte en effectieve manier volbracht worden? Ik zou mijzelf in deze situatie omschrijven als ...

Gebruik zoveel mogelijk bijvoeglijk naamwoorden en leg uit waarom die op u van toepassing zijn.

Vraag 2:

U krijgt speciale verantwoordelijkheden of opdrachten/projecten toegewezen die niet onderdeel zijn van uw routine. Hoe gaat u hiermee om? Ik zou mijzelf in deze situatie omschrijven als ...

Gebruik zoveel mogelijk bijvoeglijk naamwoorden en leg uit

waarom die op u van toepassing zijn.

Vraag 3:

U nadert een deadline en uw werkzaamheden voor een opdracht/project zijn nog niet naar wens. Hoe gaat u hiermee om? Ik zou mijzelf in deze situatie omschrijven als ...

Gebruik zoveel mogelijk bijvoeglijk naamwoorden en leg uit

waarom die op u van toepassing zijn.

Vraag 4:

U krijgt een opdracht of project toegewezen waarbij u zowel aan de wensen van de klant, als de wensen van uw collega's en leidinggevenden moet voldoen. Hoe gaat u hiermee om? Ik zou mijzelf in deze situatie omschrijven als...

Gebruik zoveel mogelijk bijvoeglijk naamwoorden en leg uit

waarom die op u van toepassing zijn.

Vraag 5:

Voelt u zich prettiger in een georganiseerde en doelgerichte omgeving of in een flexibele en ongestructureerde omgeving? Geef een voorbeeld van een situatie waaruit dat blijkt.

CORE SKILLS ASSESSMENT				
Select and comment on a minimum of three Strengths and a minimum of three Development Opportunities				
CORE SKILLS	Not Rated	Strength	Development Opportunity	Comments
Accountability	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Adaptability	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Business Acumen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Communication	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Customer Focus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Decision Making	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Developing Talent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Drive for Results	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Global Mindset	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Innovation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Strategic Agility	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Figuur 2. Performance Assessment van Fluor Corporation

Bijlage 5. Confliction Value

Tabel 5.

Gemiddelde scores van de Confliction Value.

	M	SD
Spraaktaak 1 Consciëntieusheid	.05	.17
Spraaktaak 2 Consciëntieusheid	.07	.14
Spraaktaak 3 Consciëntieusheid	.09	.23
Spraaktaak 4 Consciëntieusheid	.18	.36
Spraaktaak 5 Consciëntieusheid	.12	.26

Note. $N=90$. Minimale score is 0, maximale score is 1.