

# VU Research Portal

## Theoretische en praktische overwegingen bij spelgebaseerde en traditionele metingen van integriteit

Barends, Ard J.; de Vries, Reinout E.; van Vugt, Mark

### **published in**

Gedrag en Organisatie  
2023

### **DOI (link to publisher)**

[10.5117/GO2023.2.001.BARE](https://doi.org/10.5117/GO2023.2.001.BARE)

### **document version**

Publisher's PDF, also known as Version of record

### **document license**

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Barends, A. J., de Vries, R. E., & van Vugt, M. (2023). Theoretische en praktische overwegingen bij spelgebaseerde en traditionele metingen van integriteit. *Gedrag en Organisatie*, 36(2), 93-134.  
<https://doi.org/10.5117/GO2023.2.001.BARE>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

# Theoretische en praktische overwegingen bij spelgebaseerde en traditionele metingen van integriteit

Ard J. Barends  
*Universiteit Leiden & Vrije Universiteit Amsterdam*  
a.j.barends@law.leidenuniv.nl

Reinout E. de Vries  
*Vrije Universiteit Amsterdam*  
re.de.vries@vu.nl

Mark van Vugt  
*Vrije Universiteit Amsterdam*  
m.van.vugt@vu.nl

## Samenvatting

Integriteit is een belangrijke voorspeller voor werkgedrag, maar wordt nog weinig meegenomen in selectiesituaties. In vier studies onderzoeken we hoe de persoonlijkheidstrekk integriteit (honesty-humility) van het HEXACO persoonlijkheidsmodel gemeten kan worden met traditionele en spelgebaseerde methoden (een gegamificeerd instrument en een assessmentspel) voor selectiesituaties. In Studie 1 tonen we aan dat de vertaling van honesty-humility naar integriteit gerechtvaardigd is, aangezien het concepten zoals rechtvaardigheid en eerlijkheid meet die met morele integriteit worden geassocieerd. In Studie 2 tonen we aan dat macht van belang is voor de trekactivatie van integriteit. Dit onderzoek suggereert dan ook dat een integriteitsmeting vooral van belang is voor posities waarin kandidaten veel macht krijgen. In Studies 3 en 4 worden spelgebaseerde integriteitsmetingen gevalideerd. Deze studies ondersteunen de constructvaliditeit van de spelgebaseerde integriteitsmeting doordat deze convergeert met zelfgerapporteerde integriteit. Studie 3 laat wel zien dat een gegamificeerde integriteitsmeting net zo makkelijk te faken is als een persoonlijkheidsvragenlijst. Studie 4 laat zien dat een assessmentspel waarin deze gegamificeerde integriteitsmeting

was opgenomen, voorspellende validiteit en incrementele validiteit heeft ten opzichte van zelfgerapporteerde persoonlijkheid voor relevante uitkomsten. De resultaten tonen aan dat het mogelijk is om integriteit betrouwbaar en valide te meten met spelgebaseerde assessments en dat een spelgebaseerde integriteitsmeting voor personeelsselectie belangrijke werkgerelateerde uitkomsten kan voorspellen.

### Abstract

*Theoretical and practical considerations of using traditional and game-based assessments of honesty-humility*

Honesty-humility is an important predictor for work behavior. However, this trait is rarely taken into consideration in personnel selection. In four studies, we investigate how honesty-humility can be measured using traditional and game-based methods (a gamified instrument and an assessment game) for personnel selection. In Study 1, we demonstrate that equating honesty-humility to integrity is justified as honesty-humility covers concepts generally associated with moral integrity. In Study 2, we show that power is important for the trait activation of honesty-humility. This study suggests that honesty-humility assessments are especially important for positions in which candidates are given a lot of power. Studies 3 and 4 describe the development and validation of game-based honesty-humility assessments. These studies support the construct validity of these game-based assessments based on convergence with self-reported honesty-humility. However, Study 3 shows that a gamified assessment is just as fakeable as self-reported personality inventories. Finally, Study 4 demonstrates the predictive and incremental validity of an assessment game above and beyond self-reported personality in the prediction of relevant outcomes. The results show that it is possible to reliably and validly assess honesty-humility using game-based methods.

**Keywords:** game-based assessments, honesty-humility, integrity, personality, power

## 1 Inleiding

De persoonlijkheidstrek integriteit van het HEXACO persoonlijkheidsmodel (Ashton & Lee, 2007) is een belangrijke voorspeller van hoe men zich gedraagt op de werkvloer. Zo hebben werknemers die hoog scoren op integriteit, in

vergelijking met degenen die laag scoren op deze trek, een lagere kans om contraproductief werkgedrag te vertonen zoals diefstal, ongeoorloofd te laat komen of zich kwetsend uiten naar collega's (Anglim et al., 2018; De Vries et al., 2014; zie Pletzer et al., 2019 voor een meta-analyse). Bovendien nemen mensen die hoger scoren op integriteit, minder vaak onethische zakelijke beslissingen (Ashton & Lee, 2008) en vertonen ze ook minder vaak seksueel grensoverschrijdend gedrag (Lee et al., 2003).

Om verschillende redenen wordt er echter relatief weinig aandacht aan de persoonlijkheidstrekk integriteit besteed in personeelsselectie en assessments. De belangrijkste reden hiervoor is waarschijnlijk dat integriteit pas relatief recent is ontdekt als een losstaande persoonlijkheidstrek (Ashton et al., 2004). In de jaren negentig concludeerden persoonlijkheidspsychologen dat persoonlijkheid het beste gevat kon worden in vijf onafhankelijke persoonlijkheidstrekken, de zogenaamde Big Five (bijv. Digman, 1990; Goldberg, 1993). Recenter empirisch onderzoek naar de structuur van persoonlijkheid maakte echter duidelijk dat nog een zesde persoonlijkheidstrek, namelijk integriteit, nodig is voor een volledige beschrijving van de menselijke persoonlijkheid (Ashton et al., 2004). Integriteit blijkt ook nog eens van toegevoegde waarde voor persoonlijkheidsvragenlijsten; zo blijkt dat door de toevoeging integriteit het HEXACO model beter dan het Big Five model in staat is om uitkomsten zoals prosociaal gedrag (Thielmann et al., 2020) en contraproductief werkgedrag te voorspellen (Pletzer et al., 2019).

Integriteit lijkt dan ook een belangrijke persoonlijkheidstrek om mee te laten wegen in personeelsselectiebeslissingen. Dit artikel geeft een overzicht van een recent promotieonderzoek waarin het belang van het meten van de persoonlijkheidstrek integriteit in kaart wordt gebracht middels vragenlijsten en spelgebaseerde methoden (Barends, 2020).<sup>1</sup> In dit artikel wordt allereerst kort de achtergrond van de persoonlijkheidstrek integriteit verder besproken. Vervolgens beschrijven we hoe integriteit tot uiting komt in specifieke situaties aan de hand van het Situatie-Trek-Opbrengst-Activatie model (De Vries et al., 2016). Daarna beschrijven we hoe integriteit gemeten wordt met traditionele methoden, maar ook hoe deze trek gemeten kan worden met nieuwe spelgebaseerde methoden zoals gamificatie en assessmentgames. Tot slot geven we een overzicht van vier verschillende empirische studies.

### 1.1 Integriteit en het HEXACO persoonlijkheidsmodel

Integriteit is een onderdeel van het HEXACO persoonlijkheidsmodel (Ashton & Lee, 2007) en omvat individuele verschillen in rechtvaardigheid, oprechtheid, bescheidenheid, hebzucht en de gevoeligheid voor status.

Het HEXACO persoonlijkheidsmodel is gebaseerd op de resultaten van lexicale onderzoeken in verschillende talen naar de structuur van persoonlijkheid, waarbij alle woorden die beschrijven hoe mensen gewoonlijk denken, voelen en zich gedragen (meestal bijvoeglijke naamwoorden), zijn geanalyseerd (Ashton et al., 2004; De Raad et al., 2014; Saucier, 2009). Uit deze onderzoeken blijkt dat persoonlijkheid het beste beschreven kan worden door maximaal zes onafhankelijke dimensies. Deze zes dimensies worden samengevat in het HEXACO model, hetgeen een acroniem is van de Engelstalige namen voor de factoren. In het Nederlands heten deze factoren: Integriteit (Honesty-humility), emotionaliteit (Emotionality), extraversie (eXtraversion), verdraagzaamheid (Agreeableness), consciëntieusheid (Conscientiousness) en openheid voor ervaringen (Openness to experience).

## 1.2 Is HEXACO integriteit wel integriteit?

Een vraag rondom de zesde persoonlijkheidstrek is of de Nederlandse vertaling van honesty-humility naar integriteit wel gerechtvaardigd is, aangezien er nog veel onduidelijk is over wat er precies verstaan wordt onder integriteit (bijv. Audi & Murphy, 2006; Palanski & Yammarino, 2007; Scherkoske, 2013). Zo zijn er onderzoekers die integriteit gelijkstellen aan eerlijkheid (bijv. Butler & Cantrell, 1984), aan woord-daad consistentie (Palanski & Yammarino, 2007) of aan non-dogmatisme (bijv. Scherkoske, 2013). Opvallend is echter dat deze conclusies vooral gedaan worden op basis van theoretische argumenten. Er is in verhouding maar weinig empirisch onderzoek gedaan naar de structuur van integriteit.

Tot nu toe zijn er enkel een handvol kwalitatieve studies (bijv. Barnard, 2014; Monga, 2017; Wong et al., 2016) en twee kwantitatieve studies naar de betekenis van integriteit gedaan (Moorman et al., 2012, 2013). Uit deze studies komt ook een wisselend beeld naar voren. De belangrijkste conclusie is dat een breed scala aan concepten als onderdeel van integriteit wordt gezien, zoals moed, rechtvaardigheid en respect voor anderen.

Mogelijk nog verbazingwekkender is dat bij zowel de theoretische als de empirische verkenningen naar de betekenis van integriteit er geen gebruik wordt gemaakt van de kennis rondom de structuur van persoonlijkheid. Daarom is een van de doelen van dit artikel om meer duidelijkheid te krijgen over wat er wordt verstaan onder integriteit en hoe bijbehorende operationalisaties binnen de structuur van persoonlijkheid passen. De verwachting is dat de persoonlijkheidstrek integriteit in ieder geval een substantieel deel van de conceptualisaties van integriteit omvat. In Studie 1 onderzoeken we deze hypothese.

### 1.3 Het Situatie-Trek-Opbrengst-Activatie model

Het Situatie-Trek-Opbrengst-Activatie (STOA) model (De Vries et al., 2016) bespreekt hoe gedrag het resultaat is van de situatie en de persoon. In dit model wordt beschreven hoe personen situaties selecteren die bij hun persoonlijkheid passen. Iemand die laag op integriteit scoort, zal zich bijvoorbeeld eerder aansluiten bij een organisatie die het niet zo nauw neemt met de wet (bijv. een criminele organisatie). Integriteit is in meer algemene zin van belang in situaties waarin anderen uitgebuit kunnen worden. Het STOA model veronderstelt ook dat in verschillende situaties verschillende persoonlijkheidstreken geactiveerd worden. Zo zien we eerder verschillen in integriteit wanneer mensen een portemonnee vol geld vinden dan wanneer mensen thuis een wetenschappelijk artikel lezen. In het eerste geval wordt integriteit geactiveerd en zal iemand die laag scoort op integriteit eerder geneigd zijn het geld in eigen zak te steken en zal iemand die hoog scoort op integriteit eerder geneigd zijn het geld terug te geven. Bij het lezen van een wetenschappelijk artikel is het onwaarschijnlijk dat integriteit wordt geactiveerd. Als laatste veronderstelt het model dat de uiteindelijke opbrengsten van de uiting van een persoonlijkheidstrekk kunnen verschillen tussen situaties. Zo kan exploitatie van een ander profijt opleveren wanneer degene die geëxploiteerd wordt weinig bondgenoten heeft die wraak kunnen nemen, terwijl het uiteindelijk weinig kan opleveren wanneer degene die geëxploiteerd wordt een groot netwerk heeft dat effectief wraak kan nemen.

### 1.4 Trekactivatie van integriteit en macht

Een relevante situationele factor voor de uiting van integriteit lijkt macht te zijn, aangezien dit samenhangt met de kracht van de situatie (Cooper & Withey, 2009). Mensen in een machtspositie worden minder beperkt in hun gedrag dan mensen met beperkte macht; om die reden zijn mensen in een machtspositie dan ook in een 'zwakkere' (minder sterke of dwingende) situatie dan mensen met beperkte macht. Aangenomen wordt dat in een sterke situatie, zoals in situaties waarin mensen machteloos zijn, vooral de situatie het gedrag bepaalt en persoonlijkheid weinig effect heeft. Anderzijds wordt aangenomen dat in zwakke situaties, zoals in situaties waarin mensen veel macht hebben, vooral persoonlijkheid gedrag bepaalt. Wij verwachten dat macht dan ook van belang is voor de uiting van integriteit, aangezien personen in machtsposities veel invloed hebben op de werkcultuur en belangrijke beslissingen kunnen nemen die veel invloed hebben op anderen (Padilla et al., 2007). In Studie 2 tonen we het belang aan van integriteit in dergelijke situaties door te testen hoe variatie in macht de expressie van (lage) integriteit bij interpersoonlijke exploitatie kan inperken of versterken.

### 1.5 Traditionele en spelgebaseerde assessments

De standaardmethode om persoonlijkheid te meten is aan de hand van vragenlijsten (Rothstein & Goffin, 2006). In deze vragenlijsten beantwoorden respondenten verschillende stellingen over zichzelf. Onderzoek heeft aangetoond dat persoonlijkheidsvragenlijsten betrouwbaar en valide zijn. Alhoewel veel van de psychometrische eigenschappen behouden blijven wanneer deze vragenlijsten afgenomen worden in het kader van personeelsselectie (bijv. Anglim et al., 2017), zijn er een aantal obstakels voor het gebruik van deze vragenlijsten in personeelsselectieprocedures. Ten eerste zijn sollicitanten binnen de selectiecontext gemotiveerd om zichzelf sociaal wenselijk te presenteren en zullen dan ook tot op zekere hoogte hun antwoorden op de persoonlijkheidsvragenlijsten faken (Hu & Connelly, 2021). Onderzoek laat zien dat vooral sociaal wenselijke persoonlijkheidstreken als integriteit beïnvloed worden door sociaal wenselijke antwoorden (bijv. Anglim et al., 2017; Holtrop et al., 2021). Een tweede obstakel is dat respondenten minder positief staan tegenover persoonlijkheidsvragenlijsten in vergelijking met andere assessmentinstrumenten zoals intelligentietesten (Hausknecht et al., 2004).

Het ligt dan ook voor de hand om nieuwe instrumenten te ontwikkelen die mogelijk minder vatbaar zijn voor deze obstakels. Spelgebaseerde assessments, waarbij spelelementen gebruikt worden in assessments, lijken een mogelijke oplossing te bieden. *Spelgebaseerde assessments* is een overkoepelende term die de verschillende mogelijke toepassingen van het gebruik van spelelementen in assessments omvat: van het integreren van spelelementen in bestaande assessmentinstrumenten tot het ontwikkelen van een volwaardig computerspel om te gebruiken voor een assessment.

Momenteel zijn er echter nog weinig spelgebaseerde assessments ontwikkeld om persoonlijkheid te meten (zie voor een uitzondering Ventura & Shute, 2013). Spellen worden vooral gezien als probleemoplossingstaken en zouden daardoor beter geschikt zijn voor het meten van intelligentie (bijv. Foroughi et al., 2016; Quiroga et al., 2016). Uit onderzoek blijkt echter dat persoonlijkheid, inclusief integriteit, gerelateerd is aan gedrag in verschillende commerciële computerspellen (Tabacchi et al., 2017; Worth & Book, 2014, 2015; Zeigler-Hill & Monica, 2015). Zo blijkt bijvoorbeeld dat in het spel *World of Warcraft* integriteit negatief samenhangt met in-game gedragingen die als doel hebben om de ervaring van andere spelers te verpesten, zoals 'ganking' (het steeds weer opnieuw doden van veel zwakkere spelers; Worth & Book, 2014). Op deze manier kunnen spelers met lage integriteit hun superioriteit tegenover andere spelers laten zien. Daarnaast blijkt in het spel *Pokémon Go*, waarin spelers fysiek de straat op moeten om met de

mobiele telefoon virtuele monsters te vangen, dat extraversie positief en neuroticisme negatief samenhangen met het aantal gevangen Pokémon (Tabachhi et al., 2017). De geopperde verklaring hiervoor is dat het op straat spelen van dit spel enerzijds een sociale bezigheid is maar anderzijds ook enig sociaal ongemak met zich meebrengt. Dergelijke bevindingen laten zien dat in-game gedragingen relevante indicatoren van persoonlijkheid kunnen opleveren. Wij verwachten dan ook dat spelgebaseerde assessments daadwerkelijk ingezet kunnen worden om persoonlijkheid te meten.

In dit artikel beschrijven we de ontwikkeling en validatie van twee spelgebaseerde integriteitsassessments. Daarbij kan onderscheid worden gemaakt tussen ten minste drie soorten spelgebaseerde assessments die verschillen in de mate waarin het assessment een volwaardige spelervaring biedt, namelijk gegamificeerde assessments, assessmentgames en in-game assessments (Barends et al., 2022).

Gegamificeerde assessments hebben betrekking op het toevoegen van spelelementen aan een assessment zonder dat dit uiteindelijk in een volwaardig spel resulteert (Deterring et al., 2011). Een voordeel van een gegamificeerd assessment is dat het kan worden toegepast op een bestaand assessment en de bijbehorende psychometrische eigenschappen kan behouden, maar het uiteindelijke assessment zal waarschijnlijk meer als een assessment dan als een spel voelen (zie bijv. Landers et al., 2020). In Studie 3 beschrijven we de validering van een gegamificeerde integriteitsmeting.

Assessmentgames bieden een volwaardige spelervaring, maar bij het ontwerp wordt er wel meer gelet op het assessment en de bijbehorende psychometrie dan op de spelervaring (Barends et al., 2022). Dergelijke games kunnen dan ook bestaande assessmentitems toepassen in een spel, waarbij het spel een lineaire of adaptieve structuur volgt. In Studie 4 beschrijven we de validering van een assessmentgame om integriteit te meten.

In-game assessments zijn volwaardige computerspellen waarbij de in-game gedragingen en metingen van getoonde eigenschappen in spellen gebruikt worden voor assessments. Dit laatste type spelgebaseerde assessment is niet toegepast in ons onderzoek en laten we dan ook verder buiten beschouwing.

Spelgebaseerde assessments zouden het risico op faken kunnen verminderen doordat deelnemers van dergelijke assessments in een staat van 'flow' kunnen raken (Sweetser & Wyeth, 2005). De verwachting is daarbij dat deelnemers door het ervaren van flow zo opgaan in het spel dat zij vergeten te faken. Een aanvullend argument waarom spelgebaseerde assessments minder goed te faken zijn, is dat het mogelijkwijds minder duidelijk is voor deelnemers wat het meest wenselijke antwoord is. Naar verwachting



worden spelgebaseerde assessments ook leuker gevonden door deelnemers. Uiteraard kan het zijn dat de mate van flow, de duidelijkheid van sociaal wenselijke antwoorden en de kandidaatreacties verschillen tussen de soorten spelgebaseerde assessments, aangezien er verschillen zijn qua spelbeleving.

Hoewel het bevorderen van positieve kandidaatreacties en het tegengaan van faking mogelijk extra voordelen zijn van spelgebaseerde assessments, is het primaire doel van dit promotieonderzoek om de betrouwbaarheid en validiteit van deze instrumenten in kaart te brengen. We vergelijken dan ook maar in één van de deelprojecten in hoeverre een gegamificeerd instrument te faken is in vergelijking met een persoonlijkheidsvragenlijst (Studie 3). Kandidaatreacties zijn hier niet onderzocht. Het overgrote deel van de onderzoeken in Studies 3 en 4 richt zich dan ook op het in kaart brengen van de constructvaliditeit aan de hand van de convergente en discriminante validiteit van de assessmentgame met zelfgerapporteerde persoonlijkheid. Daarnaast wordt ook in een deelproject van Studie 4 de voorspellende validiteit van een assessmentgame onderzocht. Het is dus belangrijk om in het achterhoofd te houden dat we de voorgestelde relevantie van onze onderzoeken voor personeelsselectie vooral aantonen in 'low-stakes' situaties; we komen op dit punt terug in de algemene discussie.

## 1.6 Overzicht van de studies

In dit artikel wordt allereerst de relevantie van integriteit voor personeelsselectie in kaart gebracht. In Studie 1 onderzochten we in hoeverre de HEXACO-dimensie integriteit gelijk staat aan wat theoretisch onder integriteit verstaan wordt. In Studie 2 onderzochten we wanneer integriteit er vooral toe doet, waarbij we veronderstelden dat integriteit vooral van belang is wanneer personen zich in een machtspositie bevinden. Aan de hand van economische spellen werd gekeken of in posities met meer macht trekactivatie van integriteit plaatsvond. Vervolgens onderzochten we in Studies 3 en 4 de validiteit van spelgebaseerde metingen van integriteit. In Studie 3 stonden de ontwikkeling en validatie van een gegamificeerde integriteitsmeting centraal. In deze studie onderzochten we ook in hoeverre deze nieuwe vorm van assessment te faken was in een gesimuleerde selectiesituatie. Tot slot onderzochten we in Studie 4 de validiteit van een assessmentgame om integriteit te meten. In deze assessmentgame gebruikten we de economische spellen en gegamificeerde integriteitsmeting uit de voorgaande studies. In deze laatste studie onderzochten we ook de voorspellende validiteit van het spel en de incrementele validiteit ten opzichte van persoonlijkheidsvragenlijsten. Tabel 1 bevat een overzicht van de studies en belangrijkste bevindingen.

**Tabel 1 Schematisch overzicht van de onderzoeken, methoden en resultaten**

Studie en onderzoeksvragen	Methode	Resultaten
<p>Studie 1. Is de vertaling van honesty-humility naar integriteit te rechtvaardigen?</p>	<p>Vragenlijstonderzoek met de HEXACO-208 vragenlijst en 25 vragenlijsten van integriteitsconceptualisaties bij 416 Amerikaanse MTurkers.</p>	<p>Morele integriteitsconceptualisaties zijn vooral te plaatsen binnen de HEXACO integriteitsfactor. Moreel neutrale conceptualisaties van integriteit zijn vooral te plaatsen binnen de HEXACO consciëntieusheidsfactor.</p>
<p>Studie 2a en 2b. Wordt de trekactivatie van integriteit beïnvloed door macht?</p>	<p>Twee experimentele onderzoeken (<math>N = 209</math> en <math>N = 158</math>) waarbij de relatie tussen integriteit en exploitatief gedrag in economische spellen is vergeleken. Voor Studie 2a werd een Nederlandse gelegenheidssteekproef geworven en voor Studie 2b 158 Amerikaanse MTurkers.</p>	<p>Macht modereert de trekactivatie van integriteit in relatie tot exploitatief gedrag. Wanneer mensen veel macht hebben, dan is integriteit negatief gerelateerd aan exploitatie, wanneer macht is ingeperkt, dan is er geen relatie tussen integriteit en exploitatie.</p>
<p>Studie 3a, 3b en 3c. Is het mogelijk om met virtuele cues integriteit valide te meten? Zijn de virtuele cues moeilijker te faken dan persoonlijkheidsvragenlijsten?</p>	<p>Drie onderzoeken waarbij de convergente en discriminante validiteit van virtuele cues om integriteit te meten zijn gerelateerd aan de HEXACO persoonlijkheidsvragenlijst. Studie 3a werd uitgevoerd bij 104 Nederlandse studenten en Studie 3b bij 196 Amerikaanse MTurkers. In Studie 3c (<math>N = 191</math>) werd bij Amerikaanse MTurkers ook de fakeability van de virtuele cues vergeleken met die van de HEXACO persoonlijkheidsvragenlijst aan de hand van een gesimuleerde selectiesituatie.</p>	<p>De virtuele cues waren significant gerelateerd aan zelfgerapporteerde integriteit. De virtuele cues waren in het algemeen ongerelateerd aan de overige vijf HEXACO trekken. De virtuele cues en de HEXACO persoonlijkheidsvragenlijst waren beide in gelijke mate te faken, maar de convergente relatie met zelfgerapporteerde integriteit werd niet beïnvloed door faking.</p>
<p>Studie 4a en 4b. Is het mogelijk om met een assessmentgame integriteit valide te meten? Heeft de assessmentgame voorspellende validiteit en incrementele validiteit ten opzichte van zelfgerapporteerde persoonlijkheid in de voorspelling van relevante uitkomsten?</p>	<p>Twee onderzoeken waarbij de convergente en discriminante validiteit van de assessmentgame om integriteit te meten zijn gerelateerd aan de HEXACO persoonlijkheidsvragenlijst. Studie 4a werd uitgevoerd bij 116 Nederlanders die recent waren afgestudeerd en deelnamen aan een wedstrijd. In Studie 4b (<math>N = 241</math>) werden bij Amerikaanse MTurkers ook de voorspellende en incrementele validiteit van drie uitkomsten onderzocht.</p>	<p>De integriteitscores van de assessmentgame waren significant gerelateerd aan zelfgerapporteerde integriteit, maar ongerelateerd aan de overige vijf HEXACO trekken. De assessmentgame had voorspellende validiteit voor twee van de drie uitkomsten. Daarbij had de assessmentgame incrementele validiteit ten opzichte van de HEXACO persoonlijkheidsvragenlijsten, leertijd en geslacht in de voorspelling van valsspelen voor financieel gewin.</p>

## 2 Studie 1

We onderzochten in Studie 1 of de vertaling van de HEXACO dimensie honesty-humility, naar integriteit te rechtvaardigen is door te achterhalen in hoeverre deze persoonlijkheidstrek de kern raakt van de diverse concepten die onder integriteit verstaan worden. Uit de eerdere literatuur komt een grote variatie aan concepten naar voren die volgens de auteurs onder integriteit vallen. Bijvoorbeeld de consistentie tussen woord en daad (Palanski & Yammarino, 2007), eerlijkheid (Butler & Cantrell, 1984) en non-dogmatisme (Scherkoske, 2013). De eerdere onderzoeken hebben echter nog nooit de samenhang tussen de verschillende integriteitsconceptualisaties in kaart gebracht of binnen de reeds uitgewerkte structuur van persoonlijkheid geplaatst. Het doel van Studie 1 is daarom om te achterhalen hoe de integriteitsconcepten binnen de structuur van het HEXACO persoonlijkheidsmodel (Ashton & Lee, 2007) passen.

Na een uitvoerige inventarisatie van de literatuur rondom integriteit identificeerden we 16 integriteitsconceptualisaties met daarbij 24 unieke eigenschappen (zie Bijlage 1). We verdeelden deze integriteitsconceptualisaties onder in een morele integriteitsfactor die we moraliteit noemden en een moreel neutrale integriteitsfactor die we consistentie noemden. Onze eerste hypothese was dat de ‘moraliteit’ integriteitsfactor (bijv. eerlijkheid) samenhangt met HEXACO integriteit. Onze tweede hypothese was dat de ‘consistentie’ integriteitsfactor (bijv. de consistentie tussen woorden en daden) samenhangt met consciëntieusheid. We laten de optie open dat er nog een derde integriteitsfactor bestaat die gerelateerd is aan moed en die verband heeft met HEXACO emotionaaliteit en/of extraversie. Bovendien toont Bijlage 1 dat een paar integriteitsconceptualisaties niet passen binnen de veronderstelde twee-factorenstructuur (moraliteit en consistentie) en dat één integriteitsconceptualisatie (heelheid) niet te operationaliseren was (d.w.z. niet gemeten kon worden met behulp van een variabele).

### 2.1 Methode Studie 1

Alle studies zijn goedgekeurd door de vaste commissie wetenschappelijke ethiek van de Vrije Universiteit Amsterdam. In elke studie werden deelnemers geïnformeerd over het doel en de inhoud van het betreffende onderzoek en gaven zij middels informed consent toestemming om hun data geanonimiseerd te gebruiken voor wetenschappelijk onderzoek. Na afloop van de studie werden de deelnemers nog eens gedebriefd.

Aan Studie 1 namen 515 Amerikaanse Amazon Mechanical (MTurk) gebruikers deel. We besloten de data van 99 respondenten niet mee te nemen,

aangezien zij *noncompliance* ('*noncompliant responding*', d.w.z. het niet naleven van hoe de vragenlijst ingevuld zou moeten worden) vertoonden (Barends & De Vries, 2019). We analyseerden de data van de uiteindelijke steekproef van 416 respondenten ( $M_{\text{leeftijd}} = 40.26$  jaar,  $SD_{\text{leeftijd}} = 11.93$  jaar; 259 vrouwen). Deze respondenten vulden de Engelstalige HEXACO-208 in (De Vries et al., 2015). Daarnaast vulden zij meerdere vragenlijsten in die de 23 integriteitsconceptualisaties beoogden te meten (zie Bijlage 1) evenals de mate waarin personen morele emoties als schaamte en schuld ervaarden (Cohen et al., 2011). Meer informatie over deze studie is te vinden in Barends (2020).

## 2.2 Resultaten en discussie Studie 1

Aan de hand van een extensieanalyse (Gorsuch, 1997) op de HEXACO facetscores draaiden we eerst een exploratieve factoranalyse waaruit we zes varimax geroteerde factoren forceerden. Zoals verwacht werd de HEXACO structuur teruggevonden in deze analyse. Met de extensieanalyse werden de 25 integriteitsconceptualisaties binnen deze HEXACO factorstructuur geplaatst zonder dat zij zelf de dimensionele ruimte bepaalden. Uit de extensieanalyse bleek dat het grootste deel van de integriteitsconceptualisaties binnen de HEXACO integriteit en/of consciëntieusheid factoren te lokaliseren waren (zie Bijlage 2). Bovendien bleek dat moed het sterkste gerelateerd was aan extraversie, maar ook aan consciëntieusheid. Studie 1 ondersteunde dan ook onze voorspelling dat HEXACO integriteit de moraliteitconceptualisaties van integriteit omvat. Wel blijkt dat de consistentieconceptualisaties van integriteit sterker gerelateerd zijn aan consciëntieusheid dan aan HEXACO integriteit.

## 3 Studie 2

In Studie 2 onderzochten we de invloed van macht op trekactivatie van integriteit. Studie 2 is een replicatie en uitbreiding van het onderzoek van Hilbig en Zettler (2009). Hilbig en Zettler vonden in hun onderzoek sterkere samenhang tussen integriteit en exploitatie in het dictators spel (een zwakke situatie) dan in het ultimatumspel (een sterke situatie). Studie 2a is een directe replicatie van Hilbig en Zettler (2009) en Studie 2b is een verdere uitbreiding. In Studie 2b voegden we nog een derde economisch spel toe, het zogeheten deltaspel (Suleiman, 1996). In het deltaspel valt de machtsverhouding tussen de verdeler en ontvanger in het midden tussen het ultimatum- en dictators spel en kan dan ook worden gezien als een middelsterke situatie.

Aan de hand van het deltaspel konden we de voorspellingen vanuit het STOA model vergelijken met een contrasterende voorspelling vanuit het Traits as Situational Sensitivities (TASS) model (Marshall & Brown, 2006). Het STOA model voorspelt een lineair effect tussen macht en trekactivatie van integriteit, terwijl het TASS model veronderstelt dat de sterkste trekactivatie bij een middelsterke machtsverhouding plaatsvindt. Het STOA model neemt aan dat de trekactivatie steeds sterker wordt met de toename van de machtsverschillen omdat de situatie dan geleidelijk van een sterke situatie in een zwakke situatie overgaat en navenant ook de trekactivatie sterker wordt. Het TASS model is ontwikkeld in de context van agressie en veronderstelt dat bij een zwakke provocatie (een zwakke situatie) er geen individuele verschillen in agressie naar voren komen omdat vrijwel niemand agressief zal reageren. In een sterke situatie (een sterke provocatie) zal iedereen daarentegen agressief reageren, terwijl in een middelsterke situatie (een middelsterke provocatie) de individuele verschillen tussen wie wel en niet agressief reageren, duidelijker zichtbaar worden. Volgens het TASS model zouden daarom bij de middelsterke machtsverhouding de grootste verschillen tussen personen gevonden moeten worden.<sup>2</sup>

### 3.1 Methode Studie 2

In totaal namen 210 personen deel aan Studie 2a. De dataset van Studie 2a bestond uit een groep eerstejaars psychologiestudenten die deelnamen aan het onderzoek voor onderzoekscredits (67.5% van de steekproef) en een gelegenheidssteekproef (32.5% van de steekproef). De gelegenheidssteekproef was gerekruteerd door vier derdejaars psychologiestudenten binnen hun eigen sociale netwerken. De data van één deelnemer werden niet gebruikt, aangezien deze persoon noncompliance vertoonde op de HEXACO persoonlijkheidsvragenlijst (Barends & De Vries, 2019). De overgebleven 209 respondenten ( $M_{\text{leeftijd}} = 24.54$  jaar,  $SD_{\text{leeftijd}} = 11.42$  jaar; 151 vrouwen) vulden allereerst de Nederlandse HEXACO-100 in (De Vries et al., 2009) en speelden daarna 11 hypothetische economische spellen. Onder hypothetisch wordt verstaan dat er geen daadwerkelijke interactie met andere personen plaatsvond en dat er ook geen financiële consequenties verbonden waren aan de keuzes in het spel. Bij Studie 2a analyseerden we alleen de resultaten van de ‘verdeler’ in het ultimatumspel (Güth et al., 1982) en het dictatorspel (Forsythe et al., 1994). In beide spellen werd er gebruik gemaakt van een verdeler en een ontvanger. De verdeler bepaalde hoe een som geld tussen zichzelf en de ontvanger verdeeld werd. In het dictatorspel had de ontvanger geen zeggenschap over deze verdeling. In het ultimatumspel kon de ontvanger echter besluiten het aanbod te accepteren of te weigeren. Wanneer het

aanbod werd geaccepteerd, dan werd het geld verdeeld zoals bepaald door de verdeler. Indien de ontvanger het bod afwees, kregen de ontvanger en de verdeler niets. Meer informatie over de methode en overige informatie van dit onderzoek is te vinden in Barends et al. (2019b).

In Studie 2b verbeterden we Studie 2a op een paar vlakken. We voegden het deltaspel (Suleiman, 1996) toe als een extra economisch spel waarin er een gemiddelde machtsverhouding was tussen de verdeler en de ontvanger. In het deltaspel vermindert een weigering van de ontvanger het geboden gedrag met de helft in plaats van dit tot nul te reduceren zoals in het ultimatumspel. Studie 2b werd op twee tijdstippen uitgevoerd ( $t_1$  en  $t_2$ ). Tijdens  $t_1$  rekruteerden we 450 Amerikaanse MTurk gebruikers om de Engelstalige HEXACO-100 (Lee & Ashton, 2018) in te vullen. We nodigden voor  $t_2$  enkel de respondenten uit die geen noncompliance vertoonden (Barends & De Vries, 2019). In totaal namen 158 Amerikaanse MTurk gebruikers deel aan de  $t_2$  meting ( $M_{\text{leeftijd}} = 37.70$  jaar,  $SD_{\text{leeftijd}} = 11.46$  jaar; 80 vrouwen). In groepjes van acht respondenten speelden zij de drie economische spellen via het Software Platform for Human Interaction Experiments (SoPHIE; Hendriks, 2012). Elke respondent speelde elk spel viermaal als verdeler en viermaal als ontvanger, waarbij we de gemiddelden over de vier spellen gebruikten in de analyses. In Studie 2b vond er daadwerkelijke interactie plaats en hadden alle keuzes financiële consequenties, aangezien de bonusbetaling na deelname aan het onderzoek gebaseerd was op de hoeveelheid geld die respondenten hadden verkregen in de economische spellen (zowel in de verdeler- als in de ontvanger-rol). Zie Barends et al. (2019b) voor verdere details.

### 3.2 Resultaten en discussie Studie 2

De beschrijvende statistieken van Studie 2a en 2b zijn weergegeven in Bijlage 3 en 4.

Studies 2a en 2b repliceerden de resultaten van Hilbig en Zettler (2009). We vonden de voorspelde interactie tussen macht en integriteit (Studie 2a:  $F(1, 207) = 6.73, p = .010, \eta_p^2 = .03$ ; Studie 2b:  $F(1.25, 194.94) = 8.08, p = .003, \eta_p^2 = .05$ ; Greenhouse-Geisser gecorrigeerd). Dat wil zeggen, we vonden dat integriteit negatief samenhang met de mate van exploitatie wanneer deelnemers het dictatorspel speelden terwijl in het ultimatumspel er geen relatie was tussen integriteit en de mate van exploitatie. In tegenstelling tot de verwachtingen was er in het deltaspel ook geen relatie tussen integriteit en exploitatie. Het TASS model (Marshall & Brown, 2006) voorspelde de sterkste trekactivatie in dit spel, en het STOA model (De Vries et al., 2016) voorspelde een trekactivatie die tussen die van het ultimatumspel en het dictatorspel zou plaatsvinden. Deze resultaten waren echter tegenstrijdig met de voorspelling van beide

modellen, aangezien de relatie tussen persoonlijkheid en exploitatie in deze middelste machtspositie niet te onderscheiden was van die in de sterke situatie (het ultimatumspel).

## 4 Studie 3

Het doel van Studie 3 was om een gegamificeerde integriteitsmeting te ontwikkelen. We baseerden ons bij de ontwikkeling van de gegamificeerde integriteitsmeting op mogelijke persoonlijkheidsclues bij de inrichting van een huis of kantoor (Gosling et al., 2002) en bij de keuze voor kleding en make-up (Naumann et al., 2009; Vazire et al., 2008). Deze gegamificeerde persoonlijkheidsclues noemen we ‘virtual behavior cues’ of, in het kort, virtuele cues. Bij virtuele cues maakt de gebruiker keuzes in de vormgeving van digitale voorwerpen of omgevingen zoals het samenstellen van een avatar, het aanpassen van een virtuele auto of het inrichten van een virtueel kantoor. Eerder onderzoek heeft laten zien dat virtuele cues bij het maken van een avatar gerelateerd zijn aan Big Five persoonlijkheid (bijv. Fong & Mar, 2015). Tot nu toe is er echter nog geen onderzoek gedaan naar persoonlijkheidsclues van integriteit.

Het doel van Studies 3a en 3b was dan ook om na te gaan of het mogelijk was om dergelijke virtuele cues te ontwikkelen om integriteit te meten. Onze hypothese was dat virtuele cues van integriteit positief en significant samenhangen met zelfgerapporteerde integriteit. Bovendien onderzochten we de discriminante validiteit van deze virtuele cues van integriteit ten opzichte van de overige vijf persoonlijkheidstrekken.

Het doel van Studie 3c was vervolgens om na te gaan in hoeverre deelnemers hun respons op de virtuele cues om integriteit te meten kunnen faken. Onze eerste hypothese was dat reacties op virtuele cues minder makkelijk te faken zijn dan op zelfgerapporteerde persoonlijkheidsvragenlijsten. De reden hiervoor is dat de virtuele cues gebaseerd zijn op persoonlijkheidsclues die mogelijk onopvallender, en dus minder sociaal wenselijk, zijn dan standaard items in persoonlijkheidsvragenlijsten. Onze tweede hypothese was derhalve dat de virtuele cues van integriteit constructvaliditeit behouden zelfs wanneer respondenten proberen te faken.

### 4.1 Methode Studie 3

Aan Studie 3a namen 104 Nederlandstalige studenten deel ( $M_{\text{leeftijd}} = 21.30$  jaar,  $SD_{\text{leeftijd}} = 3.93$  jaar; 70 vrouwen). Zij vulden de Nederlandse HEXACO-100 in (De Vries et al., 2009) en maakten keuzes uit een vijftal

verschillende virtuele cues (bijv. kantoor, avatar, auto) als onderdeel van een groter onderzoek. Bij elke virtuele cue werden verschillende antwoordopties uitgewerkt op een continuüm van lage tot hoge integriteit (bijv. een kleine, goedkope auto tot een luxe sportwagen). De validatie van deze items diende dan ook om na te gaan of het mogelijk was om integriteit te meten aan de hand van dergelijke virtuele cues.

Aan Studie 3b namen 196 Amerikaanse MTurk gebruikers deel ( $M_{\text{leeftijd}} = 38.56$  jaar,  $SD_{\text{leeftijd}} = 11.41$  jaar; 86 vrouwen). Aan de hand van de virtuele cues in Studie 3a werd een set van 36 virtuele cues geconstrueerd met vier response opties die elk indicatief zijn voor een verschillende mate van integriteit. Op basis van Studie 3b werd deze set verkort tot 18 virtuele cues. Deelnemers vulden ook de Engelse HEXACO-100 in (Lee & Ashton, 2018).

Aan Studie 3c namen 191 Amerikaanse MTurk gebruikers deel ( $M_{\text{leeftijd}} = 40.83$  jaar,  $SD_{\text{leeftijd}} = 11.85$  jaar; 113 vrouwen). Zij vulden twee instrumenten in, namelijk de 18 item virtuele-cuesschaal van Studie 3b en de HEXACO-100 (Lee & Ashton, 2018). Respondenten werden willekeurig aan één van de vier condities toegewezen. In de eerste conditie vulden de respondenten beide instrumenten in zonder verdere instructies. In de tweede en derde conditie vulden zij één van de instrumenten in binnen een gesimuleerde selectiesituatie en het andere instrument zonder verdere instructies. In de vierde conditie werden beide instrumenten beantwoord als onderdeel van een gesimuleerde selectiesituatie. De gesimuleerde selectiesituatie werd opgezet met een vacaturetekst waarin de ideale kandidaat beschreven werd als iemand met een hoge mate van integriteit (bijv. de ideale kandidaat voor de positie heeft een sterk rechtvaardigheidsgevoel). Eerder onderzoek heeft laten zien dat respondenten in een gesimuleerde selectiesituatie hun antwoorden op persoonlijkheidsvragenlijsten tot op zekere hoogte faken (Shoss & Strube, 2011). Om respondenten extra te motiveren om te faken, kreeg de beste kandidaat per instrument (dus afzonderlijk voor de virtuele cues en HEXACO-100) in de gesimuleerde selectiesituatie een bonus van \$ 25,-. In de controleconditie kreeg een willekeurige respondent per instrumenten de \$ 25,- bonus. In theorie zou een respondent dus twee keer een bonus kunnen verdienen; dit gebeurde echter niet in dit onderzoek. In Studie 3c vulden de respondenten ook een manipulatiecheck in om na te gaan in hoeverre zij probeerden te faken. Respondenten beantwoordden de vraag “Ik heb de vragenlijst eerlijk ingevuld” op een 5-punts Likertschaal (1 = *zeer mee oneens*; 5 = *zeer mee eens*). Meer informatie over deze studies is te vinden in Barends et al. (2019a).



#### 4.2 Resultaten en discussie Studie 3

De beschrijvende statistieken van Studie 3a, 3b en 3c zijn weergegeven in Bijlage 5, 6 en 7. In Studies 3a, 3b en 3c vonden we dat de virtuele-cuesschaal een niet heel hoge betrouwbaarheid had (een Cronbach's alpha van .45 in Studie 3a, van .72 in Studie 3b, van .62 in Studie 3c in de faking conditie, en van .65 in de controleconditie). Wel vonden we in Studies 3a, 3b en in de controleconditie van Studie 3c dat de virtuele-cuesschaal positief en significant samenhang met zelfgerapporteerde integriteit (alle  $r$ 's > .30 en  $p$ 's < .041). Bovendien vonden we in Studie 3c dat de virtuele-cuesschaal in de geïnstrueerde faking conditie ook significant samenhang met zelfgerapporteerde integriteit in alle condities ( $r$ 's > .33,  $p$ 's < .021). In alle drie de studies vonden we dus bewijs voor de convergente validiteit van de virtuele cues aan de hand van de convergentie met zelfgerapporteerde integriteit. Verder bleek er ook sprake van discriminante validiteit aangezien er geen consistent patroon van significante correlaties met de overige vijf HEXACO persoonlijkheidstrekken werd gevonden (Barends et al., 2019a).

Om na te gaan in hoeverre reacties op de virtuele-cuesschaal te faken waren, vergeleken we allereerst de manipulatiechecks. Uit de analyse van de manipulatiechecks bleek dat deelnemers inderdaad reacties op de virtuele cues probeerden te faken wanneer zij daartoe geïnstrueerd werden ( $M_{\text{experimentele conditie}} = 2.09$ ,  $SD = 0.78$ ;  $M_{\text{controleconditie}} = 1.88$ ;  $SD = 0.54$ ),  $t(163.35) = 2.13$ ,  $p = .034$ ;  $d = 0.31$ . Hetzelfde resultaat vonden we bij respondenten die geïnstrueerd waren hun reactie op de HEXACO persoonlijkheidsvragenlijst te faken ( $M_{\text{experimentele conditie}} = 2.21$ ,  $SD = 0.84$ ;  $M_{\text{controleconditie}} = 1.78$ ;  $SD = 0.52$ ),  $t(155.91) = 4.44$ ,  $p < .001$ ;  $d = 0.62$ .

Om de daadwerkelijke mate van faking na te gaan, voerden we twee losse ANCOVAs per instrument uit. Daarbij namen we de volgorde van afname van het instrument mee als tussenpersonenfactor en leeftijd en geslacht als covariaten, aangezien de volgorde van afname soms invloed heeft op de mate van faking (Hu & Connelly, 2021) en de geselecteerde demografische variabelen gerelateerd zijn aan integriteit (Ashton et al., 2014). Allereerst vonden we dat zelfgerapporteerde integriteit inderdaad in enige mate gefaket werd, aangezien de scores in de faking conditie ( $M = 3.81$ ;  $SD = 0.66$ ) significant hoger waren dan die in de controleconditie ( $M = 3.62$ ;  $SD = 0.67$ ),  $F(1, 185) = 5.61$ ,  $p = .019$ ,  $\eta_p^2 = .03$ . Eenzelfde resultaat werd gevonden voor de virtuele cues aangezien het gemiddelde van de virtuele cues in de faking conditie ( $M = 2.67$ ;  $SD = 0.32$ ) significant hoger was dan het gemiddelde van de virtuele cues in de controleconditie ( $M = 2.50$ ;  $SD = 0.35$ ),  $F(1, 185) = 12.84$ ,  $p < .001$ ,  $\eta_p^2 = .07$ . Om na te gaan of deze faking-effecten van elkaar verschilden, voerden we een mixed MANCOVA uit waarbij de

gestandaardiseerde scores van de instrumenten als binnenpersonenfactoren werden meegenomen en de experimentele instructies en volgorde van afname als tussenpersonenfactoren. Uit deze analyse bleek dat de effecten van het faken niet verschilden tussen de twee instrumenten,  $F(1, 181) = 2.36$ ,  $p = .126$ ,  $\eta_p^2 = .01$ .<sup>3</sup>

Alhoewel het gegamificeerde assessment van integriteit aan de hand van virtuele cues te faken bleek te zijn, en hoewel de betrouwbaarheid van de virtuele cues in Studie 3 niet heel hoog was, wijzen de resultaten op de constructvaliditeit van de virtuele cues die in de assessmentgame werden aangeboden. Deze virtuele cues lijken dan ook bruikbaar om in assessmentgames te worden gebruikt voor de meting van integriteit.

## 5 Studie 4

Op basis van de resultaten van Studies 2 en 3 hebben we de assessmentgame *Building Docks* ontwikkeld om integriteit te meten. In deze assessmentgame zijn verschillende soorten taken geïntegreerd, namelijk economische spellen (Thielmann et al., 2020; Zhao & Smillie, 2015), items uit situationele beoordelingstests (oftewel Situational Judgment Tests (SJTs; Oostrom et al., 2019) en virtuele cues uit Studie 3. In Studie 4a onderzochten we de convergente en discriminante validiteit van de assessmentgame in relatie tot zelfgerapporteerde persoonlijkheid en intelligentie. In Studie 4b was ons doel om de convergente en discriminante validiteit voor wat betreft zelfgerapporteerde persoonlijkheid te repliceren en de voorspellende validiteit van het spel te onderzoeken. Daarnaast testten we ook of de assessmentgame incrementele validiteit had ten opzichte van de demografische informatie en de zelfgerapporteerde persoonlijkheid. Onze eerste hypothese was dan ook dat het assessmentspel convergente validiteit heeft in relatie tot zelfgerapporteerde integriteit. Onze tweede hypothese was dat het assessmentspel voorspellende validiteit heeft voor niet-integere gedragingen.

### 5.1 Methode Studie 4

Voor Studie 4a benaderden we 1700 recent afgestudeerde studenten die deelnamen aan een wedstrijd waarbij zij gekoppeld konden worden aan grote, internationaal opererende bedrijven. De winnaar van deze wedstrijd ontving € 10.000,-. Deelnemers aan deze wedstrijd konden geheel vrijwillig de assessmentgame *Building Docks* spelen en de Nederlandse HEXACO-100 invullen (De Vries et al., 2009). Daarnaast konden zij vrijwillig een intelligentietest – ontwikkeld door LTP business psychologen – ondergaan. In

alle communicatie maakten we duidelijk dat deelname aan ons onderzoek geen invloed had op de uitkomst van de wedstrijd. In totaal konden we van 116 deelnemers ( $M_{\text{leeftijd}} = 23.48$  jaar,  $SD_{\text{leeftijd}} = 2.06$  jaar; 50 vrouwen) de *Building Docks* en HEXACO-100 data aan elkaar koppelen, van deze groep hadden we van 75 deelnemers ook toegang tot de resultaten van de intelligentietest.

In het spel *Building Docks* dienden respondenten met drie virtuele spelers een havengebied te ontwikkelen. Het geld om dit havengebied te ontwikkelen verdienden zij aan de hand van verschillende economische spellen (bijv. het dictatorspel zoals beschreven in Studie 2). De virtuele cues uit Studie 3 werden daarbij gebruikt als beloningen voor de speler om het havengebied vorm te geven (bijv. een simpel kantoorgebouw tot een luxe kantoorgebouw). Als laatste werden de SJTs gebruikt als een afzonderlijke verhaallijn in een spel waarin de speler moest proberen een evenement te organiseren in het havengebied. Bij de SJT items werd ook elke antwoordoptie omschreven om indicatief te zijn voor verschillende niveaus van integriteit (Ostrom et al., 2019). Aangezien deze eerdere onderzoeken hebben aangetoond dat met deze deeltaken integriteit gemeten kan worden, combineerden we de resultaten van verschillende deeltaken in een totaalscore. Aangezien alle deeltaken met andere soorten responseschalen werken, werden de scores omgezet naar *z*-scores en daarna gemiddeld in een *Building Docks* integriteitsscore. Meer informatie over de studie en het spel is te vinden in Barends et al. (2022).

Voor Studie 4b verzamelden we de data van 500 Amerikaanse MTurk respondenten op drie verschillende tijdstippen ( $t_1$ ,  $t_2$  en  $t_3$ , elk met een week interval tussen de tijdstippen). 250 respondenten speelden *Building Docks* tijdens  $t_1$  en werden uitgenodigd om de HEXACO-208 (De Vries et al., 2015) tijdens  $t_2$  in te vullen. De andere 250 respondenten maakten de HEXACO-208 tijdens  $t_1$  en werden uitgenodigd om *Building Docks* te spelen tijdens  $t_2$ . Van 287 respondenten ( $M_{\text{leeftijd}} = 39.85$  jaar,  $SD_{\text{leeftijd}} = 11.59$  jaar; 170 vrouwen) hadden we zowel de *Building Docks* en HEXACO-208 data (na uitsluiting van respondenten die noncompliance vertoonden; Barends & De Vries, 2019). De data van deze 287 respondenten werden gebruikt om de constructvaliditeit van *Building Docks* te onderzoeken. Deze 287 respondenten werden ook uitgenodigd voor  $t_3$ , waarin we de drie uitkomsttaken afnamen: contraproductief werkgedrag (Bennett & Robinson, 2000), onethische zakelijke beslissingen (Ashton & Lee, 2008) en een valsspeeltaak (Moshagen & Hilbig, 2017). Onethische zakelijke beslissingen werden gemeten aan de hand van vier scenario's waarin de respondent kon aangeven in hoeverre zij wel of niet bepaalde beslissingen

zouden nemen (bijv. seksuele diensten aanbieden aan een leidinggevende om een promotie te krijgen). Voor de valsspeeltaak dienden deelnemers twee keer met een munt te gooien en kregen zij een extra financiële bonus voor de deelname aan het onderzoek als zij aangaven twee keer kop te hebben gegooid. Met deze valsspeeltaak is het dus niet te achterhalen of iemand wel of niet valsspeelt, maar kunnen we wel de prevalentie van valsspelen in de steekproef achterhalen, aangezien we verwachten dat 25% van de respondenten twee keer kop zou gooien. Deze 25% kans wordt gezien als een goede balans om voldoende onderscheidend te zijn tussen valsspelers en niet-valsspelers en de eventuele angst van deelnemers weg te nemen dat een onderzoeker op individueel niveau kan achterhalen wie er valsspeelt (Moshagen & Hilbig, 2017). Van 241 respondenten ( $M_{\text{leeftijd}} = 40.02$  jaar,  $SD_{\text{leeftijd}} = 11.35$  jaar; 146 vrouwen) hadden we ook de t3 data. Op basis van deze subgroep analyseerden we de voorspellende validiteit van *Building Docks*. Zie Barends et al. (2022) voor meer informatie.

## 5.2 Resultaten en discussie Studie 4

De resultaten van Studies 4a en 4b toonden aan dat de verschillende deeltaken alle positief en significant samenhangen (zie Bijlage 8 en 9). In beide steekproeven was de totale *Building Docks* integriteitsscore dan ook betrouwbaar (een Cronbach's alpha van .78 in Studie 4a en van .73 in Studie 4b). Bovendien was er in beide steekproeven sprake van de verwachte convergente en discriminante validiteit van de assessmentgame (zie Bijlage 8 en 9 voor de resultaten van de deeltaken). Specifiek vonden we dat in zowel Studie 4a als Studie 4b de *Building Docks* integriteitsscore positief en significant samenhang met zelfgerapporteerde integriteit (respectievelijk  $r = .33, p < .001$  en  $r = .28, p < .001$ ). Bovendien werd enkel in Studie 4b een positieve correlatie gevonden tussen de *Building Docks* integriteitsscore en een andere persoonlijkheidstrekk, namelijk verdraagzaamheid ( $r = .12, p = .036$ ). Verdraagzaamheid is de trek die in het algemeen ook het meest samenhangt met zelfgerapporteerde integriteit (bijv. De Vries et al., 2009). In beide studies waren de correlaties tussen de *Building Docks* integriteitsscore en de overige persoonlijkheidstrekken niet significant (Studie 4a alle  $[r]'s < .10, p's > .28$ ; Studie 4b alle  $[r]'s < .11, p's > .07$ ). Bovendien vonden we ook geen significante samenhang tussen de *Building Docks* integriteitsscore en de resultaten van de intelligentietest ( $r = .06, p = .633$ ). Deze resultaten laten zien dat *Building Docks* zowel een bruikbare mate van convergente als discriminante validiteit bezit die vergelijkbaar is met andere alternatieve metingen van integriteit zoals SJTs (Oostrom et al., 2019) en de virtuele cues uit Studie 3. We concluderen dan ook dat *Building Docks* in staat is

om betrouwbaar en valide integriteit te meten zonder noemenswaardige contaminatie van andere persoonlijkheidstrekken of intelligentie.

De data van Studie 4b werden ook gebruikt om de voorspellende waarde van *Building Docks* en de incrementele validiteit ten opzichte van zelfgerapporteerde persoonlijkheid en demografische variabelen te onderzoeken. De *Building Docks* integriteitsscore had voorspellende validiteit bij de valsspeeltaak ( $r = -.19, p = .004$ ) en bij onethische zakelijke beslissingen ( $r = -.21, p = .001$ ). De *Building Docks* integriteitsscore hing niet significant samen met contraproductief werkgedrag ( $r = -.12, p = .070$ ). Voor contraproductief werkgedrag vonden we wel dat de interpersoonlijke-wangedragsub-schaal significant samenhang met de *Building Docks* integriteitsscore ( $r = -.15, p = .020$ ), terwijl de organisatiegerichte-wangedragsub-schaal niet significant samenhang met de *Building Docks* integriteitsscore ( $r = -.08, p = .193$ ).

Om de incrementele validiteit van *Building Docks* te onderzoeken vergeleken we in hoeverre de *Building Docks* integriteitsscore in staat was om de uitkomsten te voorspellen ten opzichte van een model met daarin leeftijd, geslacht en alle zes zelfgerapporteerde HEXACO trekken. Op deze wijze konden we achterhalen in hoeverre *Building Docks* toegevoegde waarde heeft ten opzichte van informatie die normaal gesproken bij de afname van een persoonlijkheidsvragenlijst wordt verkregen. Voor de valsspeeltaak was het model met geslacht, leeftijd en alle zelfgerapporteerde HEXACO scores in staat om 8% van de variantie te verklaren. Het toevoegen van de *Building Docks* integriteitsscore aan dit model verbeterde de voorspelling van de valsspeeltaak significant (Nagelkerke  $R^2 = .11, \Delta R^2 = .02, p = .036$ ). Voor contraproductief werkgedrag en onethische beslissingen kon het model met geslacht, leeftijd en alle zelfgerapporteerde HEXACO scores respectievelijk 31% en 39% van de variantie verklaren. Het toevoegen van de *Building Docks* integriteitsscore aan de modellen leverde geen significante verbetering op bij de voorspelling van deze uitkomsten (respectievelijk  $R^2 = .31, \Delta R^2 = .00, p = .344$  en  $R^2 = .39, \Delta R^2 = .00, p = .323$ ).<sup>4</sup>

Deze resultaten tonen aan dat *Building Docks* een aantal relevante uitkomsten kan voorspellen en enige toegevoegde waarde heeft in de voorspelling van dergelijke uitkomsten ten opzichte van zelfrapportages. Onze bevindingen laten zien dat zelfgerapporteerde en gedragsmetingen van integriteit elk unieke voorspellende waarde hebben, hetgeen in lijn is met eerder onderzoek dat aantoonde dat de beste voorspellende waarde van zelfcontrole wordt gevonden bij een combinatie van gedragstaken en een zelfrapportage-instrument omdat zo gecompenseerd wordt voor de inherente tekortkomingen van beide methoden (Sharma et al., 2014). Een voorzichtige conclusie is dan ook dat een optimale voorspellende waarde

van integriteitsmetingen bereikt kan worden wanneer zelfrapportages aangevuld worden met gedragsmetingen zoals in de assessmentgame *Building Docks*.

## 6 Algemene discussie

De vier studies van dit promotieonderzoek tonen aan dat integriteit een belangrijke persoonlijkheidstrekk is om bij sollicitanten in kaart te brengen, vooral wanneer zij in posities van macht terecht komen (bijv. als manager). Bovendien blijkt dat integriteit – zoals gemeten met de HEXACO persoonlijkheidsvragenlijst – samenvalt met wat in het algemeen onder de morele kant van integriteit wordt verstaan. De studies tonen ook aan dat met een innovatieve spelgebaseerde assessment het mogelijk is om integriteit betrouwbaar en valide te meten. De bevindingen hebben dan ook een aantal implicaties voor personeelsselectie en assessment.

### 6.1 Praktische en theoretische implicaties

Allereerst blijkt een adequate meting van integriteit belangrijk voor de voorspelling van negatief gedrag zoals interpersoonlijke exploitatie (Studie 2), valsspelen, onethische zakelijke beslissingen en contraproductief werkgedrag (Studie 4, zie ook Pletzer et al., 2019). Het is dan ook een omissie als integriteit niet wordt meegenomen in personeelsselectie; deze omissie is des te sterker bij de selectie voor posities waarin de uiteindelijke bekleders veel macht hebben, waarbij opgemerkt dient te worden dat voor dergelijke posities zelden gestandaardiseerde assessments worden afgenomen (Hogan & Kaiser, 2005). Ons onderzoek laat zien dat in een hoge machtspositie lage integriteit tot expressie komt, terwijl de expressie van lage integriteit ingeperkt wordt wanneer mensen beperktere macht hebben. Het is dan ook vanuit het perspectief van organisaties wenselijk om kandidaten te selecteren met een hoge integriteit.

Onze onderzoeken laten ook zien dat spelgebaseerde methoden integriteit valide kunnen meten. Uit Studies 3 en 4 blijkt dan ook dat dergelijke assessmentmethoden zowel constructvaliditeit als voorspellende validiteit hebben. In het bijzonder zagen we dat de assessmentgame *Building Docks* een unieke bijdrage leverde qua voorspellende validiteit ten opzichte van de HEXACO persoonlijkheidsvragenlijst en demografische informatie. Wel moet worden opgemerkt dat deze unieke voorspellende validiteit vooral duidelijk maakt dat spelgebaseerde assessments een aanvulling kunnen zijn op het bestaande assessmentinstrumentarium, in plaats van een volledige

vervanging, aangezien in ons onderzoek de HEXACO persoonlijkheidsvragenlijst in het algemeen de beste voorspeller was van de onderzochte criteria.

De resultaten van dit onderzoek suggereren dat spelgebaseerde assessments mogelijk ook ingezet kunnen worden voor 'high-stakes' selectiesituaties. Hierbij dient wel te worden opgemerkt dat de betrouwbaarheid (d.w.z. interne consistentie) in Studies 3 en 4 niet heel hoog was, en in het algemeen nog onvoldoende was voor beslissingen op individueel niveau. Daarnaast weten we nog niets over andere relevante vormen van betrouwbaarheid, zoals test-hertestbetrouwbaarheid, van dergelijke instrumenten. Bovendien bleek dat het gegamificeerde assessment net zo goed te faken was als de persoonlijkheidsvragenlijst. Toekomstig onderzoek dient dan ook na te gaan of de validiteit van spelgebaseerde assessments behouden blijft in 'high-stakes' assessments. Een veelbelovende bevinding was dat de constructvaliditeit van het gegamificeerde assessment behouden bleef in een gesimuleerde 'high-stakes' assessment (Studie 3c).

## 6.2 Sterktes, beperkingen en richting voor toekomstig onderzoek

Een sterk punt van dit onderzoek was de combinatie van vragenlijsten, economische spellen, virtuele cues en SJTs in de losse studies en de convergentie tussen zelfrapportages van integriteit en gedragsmetingen.

Een beperking van het proefschrift is dat we relatief beperkt inzicht hebben gekregen in de voorspellende validiteit van de ontwikkelde instrumenten om integriteit mee te meten. Enkel in Studie 4b onderzochten we of het assessmentspel relevante uitkomsten kon voorspellen. Een beperking van deze studie is echter dat de data in een 'low-stakes' steekproef zijn verkregen en dat we dus nog niet weten of deze validiteit ook generaliseerd kan worden naar 'high-stakes' steekproeven waarin respondenten gemotiveerd zijn om hun antwoorden te faken (Hu & Connelly, 2021). Immers, in Studie 3c vonden we dat ons gegamificeerde assessment even goed te faken was als de persoonlijkheidsvragenlijst. Wel is het mogelijk dat assessmentgames in tegenstelling tot gegamificeerde assessments minder goed te faken zijn. Ander onderzoek lijkt hiervoor evidentie te geven (Barends & De Vries, 2022). Meer onderzoek is nodig aangezien nog weinig bekend is over de mechanismen die hieraan ten grondslag liggen. Zo kan het zijn dat assessmentgames, in tegenstelling tot gegamificeerde assessments, deelnemers in een staat van flow krijgen (Sweetser & Wyeth, 2005), waardoor ze vergeten dat ze deelnemen aan een assessment. Een andere optie is dat in-game feedback het gedrag van respondenten in het spel stuurt, waardoor faking tegengegaan kan worden. Het lijkt dan ook waardevol wanneer

toekomstig onderzoek de mate van faking van verschillende spelgebaseerde assessments vergelijkt.

Bovendien is het van belang om stil te staan bij de ecologische validiteit van de verschillende studies. Zo dient te worden opgemerkt dat het merendeel van onze studies in low-stakes onderzoekscontexten zijn gedaan en dat ook in Studie 3c de gesimuleerde high-stakes selectiesituatie wezenlijke verschillen laat zien met selectiesituaties die in de praktijk plaatsvinden. Zo zijn daadwerkelijke sollicitanten vaak intrinsiek gemotiveerd om goed over te komen in de selectiesituatie, terwijl wij enkel op basis van een financiële bonus de extrinsieke motivatie stimuleerden. Een belangrijke richting voor toekomstig onderzoek is dan ook om de validiteit van de in dit proefschrift ontwikkelde spelgebaseerde assessments te achterhalen in daadwerkelijke selectiesituaties.

Een ander punt dat niet aan de orde is gekomen in het proefschrift, is of de spelgebaseerde assessments de kandidaatreacties verbeteren (Hausknecht et al., 2004). De verwachting is dat – aangezien spellen motiverender zijn – kandidaten een dergelijke assessment ook leuker vinden dan het invullen van een persoonlijkheidsvragenlijst. Recent onderzoek op dit gebied is echter niet eenduidig. Zo blijkt een gegamificeerde SJT leuker gevonden te worden dan een tekstuele SJT (Georgiou & Nikolaou, 2020), terwijl ander onderzoek vrijwel geen verschillen vond (Landers et al., 2020). Bovendien staan sollicitanten niet per se positief tegenover nieuwe vormen van assessments. Zo blijkt bijvoorbeeld dat sollicitanten negatieve reacties hebben op geautomatiseerde interviews (Langer et al., 2019). Een momenteel nog niet beantwoorde vraag is dan ook of het leuker maken van een assessment ten koste gaat van andere kandidaatreacties zoals indrukvaliditeit en waargenomen voorspellende validiteit. Het is mogelijk dat uitkomsten van spelgebaseerde assessments minder geaccepteerd worden door kandidaten wanneer zij ongeschikt worden geacht voor een positie. Meer onderzoek naar kandidaatreacties op spelgebaseerde assessments lijkt dan ook wenselijk.

### 6.3 Conclusie

Aan de hand van dit proefschrift weten we nu dat de vertaling van de HEXACO dimensie honesty-humility naar integriteit te rechtvaardigen is, aangezien honesty-humility vooral de morele kant van integriteit omvat. Daarnaast is het van belang om integriteit te meten wanneer er selectie plaatsvindt voor machtsposities. Tot slot blijkt dat het mogelijk is om met spelgebaseerde methoden integriteit betrouwbaar en valide te meten. Spelgebaseerde instrumenten lijken dan ook een bruikbare aanvulling op het bestaande assessmentinstrumentarium.



### **Praktijkbox**

Wat betekenen deze resultaten voor de praktijk?

- HEXACO integriteit omvat wat gewoonlijk wordt verstaan onder morele integriteit. Voor integriteitsmetingen kan dan ook gebruik worden gemaakt van assessments gebaseerd op het HEXACO persoonlijkheidsmodel.
- Voor een volledige meting van integriteit dient een assessment ook consciëntieusheid en extravertie mee te nemen om de moreel neutrale integriteit-concepten zoals consistentie in woord en daad en authenticiteit te meten.
- Integriteitsassessments zijn vooral van belang voor posities waarin de bekleeder veel macht heeft, aangezien integriteit in dergelijke posities vaker geactiveerd wordt.
- Met spelgebaseerde assessments, zoals gegamificeerde assessments en assessmentgames zoals *Building Docks*, is het mogelijk om integriteit betrouwbaar en valide te meten. Dergelijke assessments hebben toegevoegde voorstellende waarde ten opzichte van standaard persoonlijkheidsvragenlijsten.

## **Dankwoord**

Wij danken LTP business psychologen voor hun financiële ondersteuning bij het onderzoek. Bovendien bedanken we LTP business psychologen en Labrooms voor de hulp bij de ontwikkeling van het assessmentspel. Tot slot bedanken wij de verschillende studenten die hebben geholpen bij het verzamelen van de data.

## **Financiering**

Het onderzoek werd gefinancierd door een beurs van LTP business psychologen in de vorm van een promotieplek voor de eerste auteur aan de Vrije Universiteit Amsterdam. De promotie is succesvol afgerond in oktober 2020.

## Bijlagen

**Bijlage 1** Integriteitsconceptualisaties en -operationalisaties en voorspelde integriteitsfactor (Studie 1)

Conceptualisatie	Beschrijving	Operationalisatie	Integriteitsfactor
1. Heelheid	Veezijdige beschrijving van de hele persoon <sup>1</sup>	Geen <sup>1</sup>	Niet
2.1 Geïntegreerde zelf: Overtuigingen	Afwezigheid van ambivalentie van overtuigingen	Revised Commitment schaal (Berzonsky et al., 2013)	Consistentie
2.2 Geïntegreerde zelf: Zelfbeeld	Stabiel zelfbeeld	Self-Concept Clarity schaal (Campbell et al., 1996)	Consistentie
3.1 & 3.2 Consistentie onder tegenslag (Moed/Morele moed)	Druk weerstaan (Morele moed en/of algemene moed) <sup>2</sup>	IPIP Values In Action Courage scale & Moral Courage subschaal van de Woodward-Pury Courage scale (Goldberg et al., 2006; Peterson & Seligman, 2004; Woodward & Pury, 2007) <sup>2</sup>	Consistentie / (Moed)
3.3 Consistentie onder tegenslag (Verleiding)	Verleiding weerstaan	Brief Self-Control schaal (Tangney et al., 2004)	Consistentie
4. Schone handen	Het hebben van vaststaande 'bottom line' principes die nooit worden opgegeven	Schlenker Integrity schaal: items met hoogste indrukvaliditeit voor schone handen (Schlenker, 2008)	Consistentie / Moraliteit
5. Trouw aan zelf	Gedrag is in lijn met eigen overtuigingen	Integrated Authenticity Scale (Knoll et al., 2015)	Consistentie
6. Consistentie in woord en daad	Consistentie tussen uitgesproken waarden en gedrag	Revised Behavioral Integrity schaal (Simons et al., 2007)	Consistentie
7. Afwezigheid van onethisch gedrag	Afwezigheid van onethisch gedrag	Moral Identity Questionnaire: Moral Integrity subschaal (Black & Reynolds, 2016)	Moraliteit
8.1 Algemeen moreel besef (Superieur)	Moreel superieur gedrag vertonen	Moral Identity Questionnaire: Moral Self subschaal (Black & Reynolds, 2016)	Moraliteit
8.2 Algemeen moreel besef (Normatief)	Een normatief ethisch standpunt innemen	Ethics Position Questionnaire: Moral Idealism subschaal (Forsyth, 1980)	Moraliteit

Conceptualisatie	Beschrijving	Operationalisatie	Integriteitsfactor
9. Eerlijkheid	Eerlijk gedrag	Eerlijkheidsitems van de MACH-IV en Schlenker Integrity schaal (Christie & Geis, 1970; Schlenker, 2008)	Moraliteit
10. Betrouwbaarheid	Betrouwbaar gedrag	Positive Reciprocity schaal (Perugini et al., 2003)	Moraliteit
11.1 Rechtvaardigheid	Rechtvaardig interpersoonlijk gedrag	IPIP Values In Action Justice schaal (Goldberg et al., 2006; Peterson & Seligman, 2004)	Moraliteit
11.2 Respect	Anderen met respect behandelen	Unconditional Respect for Others schaal (Laljee et al., 2007)	Moraliteit
12.1 Openheid/Authenticiteit (Zelfbeeld)	Zelf zien als een moreel persoon	Tennessee Self-Concept schaal: Morality subschaal (Marsh & Richards, 1988)	Consistentie / Moraliteit
12.2 Openheid/Authenticiteit (Ware zelf)	Ware zelf tonen	Relational Transparency subschaal van de Authentic Leadership Scale (Northouse, 2013; Walumbwa et al., 2008)	Consistentie / Moraliteit
13.1 & 13.2 Morele doelmatigheid	Zelfreflectie op eigen moraliteit	Moral Attentiveness schaal (Reynolds, 2008) <sup>3</sup>	Niet
14. Juiste achting voor eigen oordeel	Afwijzen van moreel relativisme	Ethics Position Questionnaire: Moral Relativism schaal (Forsyth, 1980)	Niet
15. Non-dogmatisme	Overtuigingen willen bijstellen op redelijke gronden	Dogmatism Scale (Altemeyer, 2002)	Niet
16.1 Empathie	Empathisch naar anderen zijn	Interpersonal Reactivity Index (Davis, 1980; 1983)	Niet
16.2 Compassie	Compassie tonen naar anderen	The Compassion Scale (Pommier, 2010)	Niet
17.1 Morele emoties (schaamte)	Tendens om gevoelens van schaamte te ervaren	Guilt And Shame Proneness schaal (Cohen et al., 2011)	?
17.2 Morele emoties (schuld)	Tendens om gevoelens van schuld te ervaren	Guilt And Shame Proneness schaal (Cohen et al., 2011)	?

NB. <sup>1</sup> = Aangezien heelheid de algehele beschrijving van de persoon omvat, kan dit gelijk worden gesteld aan persoonlijkheid; <sup>2</sup> = Op basis van de literatuur over moed is ervoor gekozen om bij druk weerstaan onderscheid te maken tussen algemene moed en morele moed; <sup>3</sup> = Moral Attentiveness blijkt multidimensioneel te zijn, daarom is zowel de reflectie- als de perceptieschaal meegenomen.

**Bijlage 2** Extensieanalyse van de HEXACO schalen en de integriteitsconceptualisaties in Studie 1

Concept	Schaal	$\alpha$	H	E	X	A	C	O
<b>Kernvariabelen</b>								
1. H Oprechtheid	HEXACO	.85	<b>.73</b>	-.04	.05	.11	.17	.04
2. H Rechtvaardigheid	HEXACO	.85	<b>.63</b>	.22	.06	.20	<b>.34</b>	-.05
3. H Hebzuchtvermijding	HEXACO	.87	<b>.72</b>	-.06	-.14	.15	-.02	.03
4. H Bescheidenheid	HEXACO	.87	<b>.73</b>	.17	-.21	.20	.06	-.06
5. E Angstigheid	HEXACO	.79	.17	<b>.49</b>	<b>-.36</b>	-.08	-.05	-.27
6. E Bezorgdheid	HEXACO	.87	-.10	<b>.47</b>	<b>-.60</b>	<b>-.32</b>	.01	-.06
7. E Afhankelijkheid	HEXACO	.83	-.15	<b>.61</b>	-.01	-.02	-.16	-.01
8. E Sentimentaliteit	HEXACO	.82	.24	<b>.79</b>	.07	.15	.16	.10
9. X Sociale zelfwaardering	HEXACO	.88	.16	-.01	<b>.59</b>	.23	<b>.41</b>	-.01
10. X Sociale bravoure	HEXACO	.88	-.22	-.18	<b>.76</b>	.11	.21	.20
11. X Sociabiliteit	HEXACO	.85	<b>-.30</b>	.29	<b>.66</b>	.24	.11	.06
12. X Levendigheid	HEXACO	.86	.01	.04	<b>.73</b>	.28	<b>.35</b>	.08
13. A Vergevingsgezindheid	HEXACO	.87	.09	.04	<b>.33</b>	<b>.64</b>	-.05	.08
14. A Aanpassingsbereidheid	HEXACO	.75	.24	.03	.13	<b>.73</b>	.09	-.02
15. A Zachtaardigheid	HEXACO	.83	.20	.16	.14	<b>.81</b>	.07	.12
16. A Geduld	HEXACO	.82	.15	-.12	.17	<b>.73</b>	.18	.19
17. C Ordelijkheid	HEXACO	.87	.18	-.03	.13	.10	<b>.69</b>	-.05
18. C Perfectionisme	HEXACO	.77	.04	.12	.01	-.11	<b>.66</b>	.22
19. C Bedachtzaamheid	HEXACO	.81	<b>.36</b>	-.09	.09	.20	<b>.66</b>	.09
20. C IJver	HEXACO	.83	.05	-.06	<b>.34</b>	.08	<b>.76</b>	.18
21. O Esthetische waardering	HEXACO	.83	.13	.16	.07	.10	.15	<b>.73</b>
22. O Creativiteit	HEXACO	.81	-.13	-.06	.24	.11	.12	<b>.74</b>

Concept	Schaal	$\alpha$	H	E	X	A	C	O
23. O Weetgerigheid	HEXACO	.79	.01	-.12	.06	.08	.16	<b>.72</b>
24. O Onconventionaliteit	HEXACO	.70	.03	.04	-.02	.02	-.01	<b>.71</b>
25. Altruïsme	HEXACO	.80	<b>.46</b>	<b>.55</b>	.04	<b>.31</b>	.24	.21
26. Proactiviteit	HEXACO	.80	-.12	.02	<b>.46</b>	.12	<b>.64</b>	.27
<b>Integriteitsconcepten</b>								
1. Trouw aan zelf	IAS	.72	.27	-.03	<b>.33</b>	.01	<b>.48</b>	.15
2. Openheid/Authenticiteit (Ware zelf)	ALS: RA	.67	.28	.21	<b>.35</b>	.20	<b>.37</b>	.15
3. Afwezigheid van onethisch gedrag	MIQ: MI	.86	<b>.64</b>	.15	-.00	.10	<b>.35</b>	.07
4. Algemeen moreel besef (Superieur)	MIQ: MS	.84	<b>.39</b>	.29	.08	.11	<b>.43</b>	.16
5. Geïntegreerde zelf: Zelfbeeld	SSC	.89	<b>.42</b>	-.14	.23	.12	<b>.47</b>	-.04
6. Openheid/Authenticiteit (Zelfbeeld)	TSCS M	.83	<b>.34</b>	.09	.22	.18	<b>.49</b>	.04
7. Consistentie onder tegenslag (Morele moed)	WP-23: MC	.77	-.02	-.07	<b>.53</b>	.03	<b>.31</b>	.29
8. Consistentie onder tegenslag (Algemene moed)	VIA: C	.49	.24	.18	.13	.15	.16	.26
9. Rechtvaardigheid	VIA: J	.83	<b>.43</b>	.23	.17	.28	<b>.35</b>	.16
10. Schone handen	SIS	.74	<b>.36</b>	-.20	.03	-.01	<b>.37</b>	.15
11. Eerlijkheid	MACH-IV & SIS	.84	<b>.56</b>	.14	.13	.13	<b>.32</b>	-.02
12. Algemeen moreel besef (Normatief)	EPQ: MI	.80	<b>.38</b>	<b>.31</b>	.01	.16	.28	.05
13. Juiste achtung voor eigen oordeel	EPQ: MR	.80	-.18	-.03	-.00	.07	-.17	.05
14. Empathie	IRI	.83	.29	<b>.49</b>	.14	<b>.40</b>	.29	.22
15. Compassie	TCS	.83	<b>.38</b>	<b>.50</b>	.08	.25	<b>.33</b>	.17
16. Morele doelmatigheid (perceptie)	MA: MP	.76	<b>-.31</b>	.11	.15	.01	-.02	.12
17. Morele doelmatigheid (reflectie)	MA: MR	.79	.02	<b>.33</b>	.04	.04	.16	<b>.35</b>
18. Betrouwbaarheid	PRS	.84	.24	.28	.02	.14	<b>.41</b>	.14
19. Respect	URO	.79	<b>.51</b>	.22	.07	<b>.34</b>	<b>.31</b>	.16
20. Geïntegreerde zelf: Overtuigingen	R-CS	.83	.24	.08	<b>.37</b>	.15	<b>.54</b>	.09

Concept	Schaal	$\alpha$	H	E	X	A	C	O
21. Consistentie in woord en daad	R-BI	.90	.26	.13	.18	.10	.56	.13
22. Consistentie onder tegenslag (Verleiding)	BSCS	.82	.19	-.11	.33	.30	.52	.05
23. Non-dogmatisme	DOG	.74	-.06	-.03	.20	-.09	.11	-.20
24. Morele emoties (schaamte)	GASP	.35	.07	.34	-.26	-.09	.11	-.01
25. Morele emoties (schuld)	GASP	.73	.43	.31	-.07	.11	.36	.11

NB. N = 416; absolute factorladingen gelijk aan of groter dan .30 zijn vetgedrukt gerapporteerd. GASP = Guilt And Shame Proneness schaal; MIQ: I = Moral Identity Questionnaire Integrity subschaal; MIQ: MS = Moral Identity Questionnaire Moral Self subschaal; IRI = Interpersonal Reactivity Index; DOG = Dogmatism Scale; SSC = Self-Concept Clarity schaal; URO = Unconditional Respect for Others schaal; BSCS = Brief Self-Control schaal; R-BI = Revised Behavioral Integrity schaal; IAS = Integrated Authenticity Scale; R-CS = Revised-Commitment schaal; PRS = Positive Reciprocity schaal; MA: MR = Moral Attentiveness; Moral Reflectiveness schaal; MA: MP = Moral Attentiveness; Moral Perception schaal; TCS = The Compassion Scale; SIS = Schlenker Integrity schaal; ALS: RA = Authentic Leadership Scale; Relational Transparency schaal; TSCS M = Tennessee Self-Concept schaal: Morality subschaal; VIA: Courage = IPIP Values In Action Courage schaal; VIA: J = IPIP Values In Action Justice schaal; WP-23: MC = Woodward-Pury Moral Courage subschaal; MACH-IV & SIS zijn de eerlijkheidsitems van de MACH-IV schaal en de Schenkler Integrity schaal; EPQ: MI = Ethics Position Moral Idealism subschaal; Ethics Position Moral Relativism subschaal.

**Bijlage 3 Correlatietabel tussen de variabelen in Studie 2a en keuzes in de economische spellen**

Variabele	$\alpha$	M	SD	1	2	3	4	5	6	7	8	9
1. Geslacht	-	0.28	0.45	-								
2. Leeftijd	-	24.53	11.42	.17*	-							
3. H	.81	3.54	0.51	-.30**	.09	-						
4. E	.85	3.25	0.58	-.43**	-.18**	.18**	-					
5. X	.84	3.54	0.51	-.03	-.02	.02	-.17*	-				
6. A	.86	3.03	0.58	.05	-.04	-.04	.25**	-.13	.22**	-		
7. C	.84	3.47	0.54	-.00	.14*	.19**	.06	-.04	-.01	-.01	-	

Variabele	$\alpha$	M	SD	1	2	3	4	5	6	7	8	9
8. O	.83	3.21	0.60	.07	.04	.07	-.10	.08	.03	.02	-	
9. Percentage geld gehouden in het dictatorspel	-	54.89	14.57	.09	.10	-.27**	-.02	.07	-.06	.05	-.03	
10. Percentage geld gehouden in het ultimatumspel	-	51.65	9.97	-.11	-.03	-.13	.11	.01	-.13	-.10	-.04	.30**

NB. Bij geslacht V = 0 en M = 1; \* $p < .05$ ; \*\* $p < .01$ . Hogere getallen in de economische spellen zijn indicatief voor meer exploitatie.

#### Bijlage 4 Correlatietabel tussen de variabelen in Studie 2b en keuzes in de economische spellen

Variabele	$\alpha$	M	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. Geslacht	-	0.49	0.50	-											
2. Leeftijd	-	36.70	11.46	-.12	-										
3. H	.87	3.59	0.70	-.19*	.15	-									
4. E	.87	3.14	0.70	-.46**	-.05	-.08	-								
5. X	.93	3.13	0.81	-.01	.03	.03	-.17*	-							
6. A	.92	3.16	0.74	.02	.12	.39**	-.20*	.34**	-						
7. C	.89	3.84	0.60	-.16*	.04	.20*	-.12	.37**	.15	-					
8. O	.85	3.59	0.63	-.08	-.02	-.01	.01	.31**	.22*	.17*	-				
9. Percentage geld gehouden in het dictatorspel	.87	66.43	19.03	.23*	-.11	-.25**	-.09	.16*	.01	.03	-.07	-			
10. Percentage geld gehouden in het deltaspel	.87	57.76	12.15	.01	-.10	-.06	-.00	.14	.02	.07	-.10	.39**	-		
11. Percentage geld gehouden in het ultimatumspel	.90	56.59	11.53	.09	-.13	-.03	-.03	.12	.02	.02	-.17*	.29**	.81**	-	
12. Daadwerkelijk verkregen geld in het deltaspel	-	53.28	6.37	-.10	-.03	-.05	-.04	.00	.00	.07	-.08	.26**	.52**	.26**	-
13. Daadwerkelijk verkregen geld in het ultimatumspel	-	50.73	16.62	-.07	.06	-.02	-.10	-.07	-.07	.05	-.03	-.08	-.40**	-.52**	.30*

NB. Bij geslacht V = 0 en M = 1; \* $p < .05$ ; \*\* $p < .01$ . Hogere getallen in de economische spellen zijn indicatief voor meer uitbuiting.

**Bijlage 5 Correlatietabel tussen de demografische variabelen, persoonlijkheid en virtuele cues (Studie 3a)**

Variabele	M	SD	1	2	3	4	5	6	7	8
1. Geslacht	0.33	0.47	-							
2. Leeftijd	21.40	3.93	.09	-						
3. H	3.32	0.70	-.37**	.03	-					
4. E	3.14	0.66	-.51**	-.11	.26**	-				
5. X	3.55	0.60	.13	-.16	-.17	-.04	-			
6. A	3.05	0.57	-.13	.04	.47**	.16	.01	-		
7. C	3.35	0.55	-.21*	.01	.28**	.22*	-.19*	.01	-	
8. O	3.45	0.61	.15	.05	-.09	-.11	.09	-.13	-.17	-
9. Virtuele cues	0.00	0.52	-.30*	-.08	.33**	.26**	-.09	.17	-.07	.04

NB. Bij geslacht V = 0 en M = 1; \*  $p < .05$ ; \*\*  $p < .01$ .

**Bijlage 6 Correlatietabel tussen de demografische variabelen, persoonlijkheid en virtuele cues (Studie 3b)**

Variabele	M	SD	1	2	3	4	5	6	7	8	9
1. Geslacht	0.56	0.50	-								
2. Leeftijd	38.56	11.41	-.24*	-							
3. H	3.57	0.71	-.22*	.29**	-						
4. E	3.09	0.68	-.40*	.07	.00	-					
5. X	3.21	0.84	.15*	.09	.07	-.30**	-				
6. A	3.32	0.66	-.09	.09	.37**	-.17*	.41**	-			
7. C	3.85	0.52	-.06	-.08	.25**	-.17*	.38**	.24**	-		
8. O	3.56	0.71	.07	.09	.20**	-.06	.32**	.15*	.27**	-	
9. 36 item VC	2.45	0.31	-.11	.24**	.40**	.13	-.25**	.08	-.14*	.03	-
10. 18 item VC	2.48	0.39	-.13	.27**	.49**	.13	-.16*	.15*	-.11	.00	.91**

NB. Bij geslacht V = 0 en M = 1; VC = virtuele cues; \*  $p < .05$ ; \*\*  $p < .01$ .



**Bijlage 7** Correlatietabel tussen de demografische variabelen, persoonlijkheid en de virtuele cues in de controle en faking conditie (Studie 3c)

Variabele	M	SD	VC (controle)	VC (faking)
Geslacht	0.64	0.48	-.07	.17
Leeftijd	40.83	11.85	.21*	.06
H (controle)	3.62	0.67	.30*	.34*
E (controle)	3.16	0.64	-.02	.29*
X (controle)	3.43	0.63	-.09	-.01
A (controle)	3.40	0.54	.15	.19
C (controle)	3.99	0.53	-.07	-.01
O (controle)	3.67	0.63	-.26	-.04
H (faking)	3.81	0.66	.32*	.38**
E (faking)	3.08	0.61	-.21	-.11
X (faking)	3.59	0.72	-.13	-.20
A (faking)	3.46	0.67	.16	.25
C (faking)	4.19	0.51	-.01	-.04
O (faking)	3.70	0.66	-.03	.04

*NB.* Bij geslacht  $V = 0$  en  $M = 1$ ; VC = virtuele cues; \*  $p < .05$ ; \*\*  $p < .01$ .

**Bijlage 8** Correlatietabel tussen de demografische variabelen, persoonlijkheid, intelligentie en het assessmentspel (Studie 4a)

	<i>M</i>	<i>SD</i>	<i>α</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Geslacht	0.57	0.50	-																
2. Leeftijd	23.48	2.06	-	-.06															
3. H	3.67	0.44	.76	-.17	.01														
4. E	2.94	0.54	.82	-.49**	-.15	.03													
5. X	3.63	0.61	.90	-.22*	.20*	-.06	-.04												
6. A	3.31	0.49	.80	-.08	.05	.08	.09	.01											
7. C	3.71	0.52	.85	-.16	-.03	.14	.07	.12	-.10										
8. O	3.43	0.64	.85	-.04	.31**	-.01	-.00	.30**	.07	.02									
9. Altruïsme	3.84	0.59	.57	-.28**	.11	.44**	.18	.20*	.24*	.10	.21*								
10. Abstracte intel-	4.96	1.04	-	.13	-.13	-.17	-.04	-.20	.17	.23	.02	-.06							
ligentie																			
11. Num-merieke intel-	4.97	1.10	-	.30**	-.21	-.19	-.26*	.13	-.13	.26*	-.18	-.14	.37**						
ligentie																			
12. Verbale intel-	3.98	1.13	-	.22	-.01	-.13	-.08	.05	.02	.08	.03	.07	.53**	.37**					
ligentie																			
13. Totale intelli-	4.79	1.07	-	.25*	-.14	-.19	-.14	-.04	.05	.23	-.04	-.05	.85**	.67**	.82**				
gentie																			
14. BD:EG	0.00	0.42	.62	.23*	.03	.10	-.10	-.19*	-.01	.02	.08	-.02	.10	.15	.19	.18			
15. BD:SJT	0.00	0.45	.45	.07	.05	.17	.02	.04	.07	-.02	-.03	.03	-.16	-.10	-.07	-.14	.22*		
16. BD:VC	0.00	0.45	.76	-.09	.18	.43**	-.04	-.02	.16	.21*	.14	.22*	.11	.04	.05	.09	.29**	.33**	
17. BD:totale score	0.00	0.32	.78	.09	.12	.33**	-.06	-.08	.10	.10	.09	.11	.02	.04	.07	.06	.68**	.73**	.75**

NB. Bij geslacht  $V = 0$ ,  $M = 1$ ; \*  $p < .05$ ; \*\*  $p < .01$ . BD = Building Docks, EG = economisch spel, SJT = Situational Judgment Test, VC = virtuele cues.

**Bijlage 9** Correlatietabel tussen de demografische variabelen, persoonlijkheid, het assessmentspel en de uitkomstmaten (Studie 4b)

	M	SD	$\alpha$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Geslacht	0.41	0.49	-																
2. Leeftijd	39.85	11.59	-	-.17**															
3. H	3.70	0.68	.94	-.23**	.27**														
4. E	3.18	0.60	.91	-.39**	.09	.12													
5. X	3.23	0.74	.95	.07	.12*	-.09	-.24**												
6. A	3.16	0.59	.92	.05	.13*	.37**	-.02	.33**											
7. C	3.75	0.55	.91	-.04	.06	.32**	-.06	.30**	.17**										
8. O	3.41	0.58	.90	.07	.08	.13*	-.12*	.12*	.17**	.16**									
9. BD EG	0.00	0.46	.68	.05	.09	.15**	-.04	.02	.09	-.06	-.01								
10. BD SJT	0.00	0.43	.37	-.20**	.13*	.26**	.14*	-.14*	.03	.08	-.08	.13*							
11. BD VC	0.00	0.45	.75	-.02	.09	.14*	.00	-.09	.13*	-.03	.01	.14*	.19**						
12. BD totale score	0.00	0.29	.73	-.08	.16**	.28**	.05	.11	.12*	-.01	-.04	.65**	.65**	.68**					
13. UBD	1.96	0.64	.78	.22**	-.31**	-.46**	-.26**	.10	-.26**	-.10	-.26**	-.09	-.19**	-.13	-.21**				
14. CWB:interp.	1.47	0.85	.90	.15*	-.13*	-.34**	-.06	-.07	-.24**	-.34**	-.17*	-.06	-.16*	-.08	-.15*	.31**			
15. CWB org.	1.76	0.85	.87	.04	-.18**	-.35**	.05	-.31**	-.27**	-.47**	-.13*	-.05	-.09	-.03	-.08	.21**	.69**		
16. CWB totaal	1.65	0.79	.92	.09	-.17**	-.38**	.01	-.24**	-.28**	-.46*	-.15*	-.06	-.13	-.05	-.12	.27**	.87**	.96**	
17. Valspeeltaak (aantal keer kop)	0.66	0.47	-	.03	-.12	-.12	-.04	.10	-.03	.08	.09	-.18**	-.12	-.08	-.19**	.18**	.05	.00	.02

NB. Bij geslacht  $V = 0$ ,  $M = 1$ ; \*  $p < .05$ ; \*\*  $p < .01$ . BD = Building Docks; EG = economisch spel, SJT = Situational Judgment Test, VC = virtuele cue, UBD = oneethische zakelijke beslissingen, CWB = contra-productief werkgedrag; interp = interpersoonlijk; org = organisatiegericht.

## Noten

1. We laten één hoofdstuk van dit proefschrift buiten beschouwing dat minder relevant is voor de hoofdlijnen van het proefschrift zoals hier beschreven. In dit hoofdstuk worden de effecten van verschillende uitsluitingscriteria in online onderzoeken vergeleken. De geïnteresseerde lezer verwijzen we naar Barends en De Vries (2019) voor meer informatie.
2. We hebben deze TASS voorspelling meegenomen tijdens het revisieproces van Barends et al. (2019b) op basis van de suggestie van de editor. We hadden dus a priori geen verwachtingen welke kant van de integriteitsdistributie meer of minder geactiveerd zou worden in het deltaspel.
3. De resultaten zijn vrijwel vergelijkbaar wanneer we geen covariaten meenemen of controleren voor de volgorde van afname van de instrumenten. Zo bleek uit de onafhankelijke t-test dat zelfgerapporteerde integriteit in de faking conditie ( $M = 3.81$ ,  $SD = 0.66$ ) marginaal significant hoger was dan zelfgerapporteerde integriteit in de controleconditie ( $M = 3.62$ ,  $SD = 0.66$ ),  $t(189) = 1.93$ ,  $p = .056$ ,  $d = .28$ . Ook bleek uit een tweede onafhankelijke t-test dat de score op de virtuele-cuesschaal in de faking conditie ( $M = 2.67$ ,  $SD = 0.32$ ) significant hoger was dan in de controleconditie ( $M = 2.50$ ,  $SD = 0.35$ ),  $t(189) = 3.41$ ,  $p < .001$ ,  $d = .34$ . De mixed MANOVA zonder het volgorde-effect gaf desondanks aan dat de mate van faking niet significant verschilde tussen de twee instrumenten,  $F(2, 186) = 2.07$ ,  $p = .129$ ,  $\eta_p^2 = .02$ .
4. Deze resultaten rondom de incrementele validiteit van de integriteitscore van *Building Docks* blijven hetzelfde wanneer we enkel een vergelijking maken met zelfgerapporteerde integriteit. Zelfgerapporteerde integriteit voorspelt namelijk geen significante variantie in de valsspeeltaak (Nagelkerke  $R^2 = .02$ ,  $p = .052$ ), terwijl het toevoegen van de *Building Docks* integriteitsscore aan dit model tot een significante verbetering van de voorspelling van de valsspeeltaak leidt ( $\Delta R^2 = .04$ ,  $p = .009$ ).

## Literatuur

- Altemeyer, B. (2002). Dogmatic behavior among students: Testing a new measure of dogmatism. *The Journal of Social Psychology*, 142(6), 713-721. <https://doi.org/10.1080/00224540209603931>
- Anglim, J., Lievens, F., Everton, L., Grant, S. L., & Marty, A. (2018). HEXACO personality predicts counterproductive work behavior and organizational citizenship behavior in low-stakes and job applicant contexts. *Journal of Research in Personality*, 77, 11-20. <https://doi.org/10.1016/j.jrp.2018.09.003>
- Anglim, J., Morse, G., De Vries, R. E., MacCann, C., & Marty, A. (2017). Comparing job applicants to non-applicants using an item-level bifactor model on the

- HEXACO personality inventory. *European Journal of Personality*, 31(6), 669-684. <https://doi.org/10.1002/per.2120>
- Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, 11(2), 150-166. <https://doi.org/10.1177/1088868306294907>
- Ashton, M. C., & Lee, K. (2008). The prediction of honesty-humility-related criteria by the HEXACO and five-factor models of personality. *Journal of Research in Personality*, 42(5), 1216-1228. <https://doi.org/10.1016/j.jrp.2008.03.006>
- Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, 18(2), 139-152. <https://doi.org/10.1177/1088868314523838>
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., Boies, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psychological studies in seven languages. *Journal of Personality and Social Psychology*, 86(2), 356-366. <https://doi.org/10.1037/0022-3514.86.2.356>
- Audi, R., & Murphy, P. E. (2006). The many faces of integrity. *Business Ethics Quarterly*, 16(1), 3-21. <https://doi.org/10.5840/beq20061615>
- Barends, A. J. (2020). *Assessing honesty-humility: Theoretical and practical considerations of traditional and game-based assessments*. Repro VU.
- Barends, A. J., & De Vries, R. E. (2019). Noncompliant responding: Comparing exclusion criteria in MTurk personality research to improve data quality. *Personality and Individual Differences*, 143, 84-89. <https://doi.org/10.1016/j.paid.2019.02.015>
- Barends, A. J., & De Vries, R. E. (2022). Construct validity of a personality assessment game in a simulated selection situation and the moderating roles of the ability to identify criteria and dispositional insight. *International Journal of Selection and Assessment*. <https://doi.org/10.1111/ijasa.12404>
- Barends, A. J., De Vries, R. E., & Van Vugt, M. (2019a). Gamified personality assessment: Virtual behavior cues of honesty-humility. *Zeitschrift für Psychologie*, 227(3), 207-217. <https://doi.org/10.1027/2151-2604/a000379>
- Barends, A. J., De Vries, R. E., & Van Vugt, M. (2019b). Power influences the expression of honesty-humility: The power-exploitation affordances hypothesis. *Journal of Research in Personality*, 82, 103856. <https://doi.org/10.1016/j.jrp.2019.103856>
- Barends, A. J., De Vries, R. E., & Van Vugt, M. (2022). Construct and predictive validity of an assessment game to measure honesty-humility. *Assessment*, 29(4), 630-650. <https://doi.org/10.1177/1073191120985612>
- Barnard, A. (2014). The competencies of integrity. *Journal of Psychology in Africa*, 21(2), 267-273. <https://doi.org/10.1080/14330237.2011.10820455>
- Bennett, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology*, 85(3), 349-360. <https://doi.org/10.1037/0021-9010.85.3.349>

- Berzonsky, M. D., Soenens, B., Luyckx, K., Smit, I., Papini, D. R., & Goossens, L. (2013). Development and validation of the revised identity style inventory (ISI-5): Factor structure, reliability, and validity. *Psychological Assessment, 25*(3), 893-904. <https://doi.org/10.1037/a0032642>
- Black, J. E., & Reynolds, W. M. (2016). Development, reliability, and validity of the moral identity questionnaire. *Personality and Individual Differences, 97*, 120-129. <https://doi.org/10.1016/j.paid.2016.03.041>
- Butler, J. K., & Cantrell, S. (1984). A behavioral decision theory approach to modeling dyadic trust in superiors and subordinates. *Psychological Reports, 55*(1), 19-28. <https://doi.org/10.2466/pro.1984.55.1.19>
- Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology, 70*(1), 141-156. <https://doi.org/10.1037/0022-3514.70.1.141>
- Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. Academic Press.
- Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology, 100*(5), 947-966. <https://doi.org/10.1037/a0022641>
- Cooper, W. H., & Withey, M. J. (2009). The strong situation hypothesis. *Personality and Social Psychology Review, 13*(1), 62-72. <https://doi.org/10.1177/1088868308329378>
- Crowson, H. M. (2009). Does the DOG scale measure dogmatism? Another look at construct validity. *The Journal of Social Psychology, 149*(3), 265-283. <https://doi.org/10.3200/SOCP.149.3.365-383>
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126. <https://doi.org/10.1037/0022-3514.44.1.113>
- De Raad, B., Barends, D. P., Timmerman, M. E., De Rover, K., Mlačić, B., & Church, A. T. (2014). Towards a pan-cultural personality structure: Input from 11 psycholexical studies. *European Journal of Personality, 28*(5), 497-510. <https://doi.org/10.1002/per.1953>
- De Vries, A., De Vries, R. E., Born, M. Ph., & Van den Berg, R. H. (2014). Persoonlijkheid als voorspeller van werkprestatie en contraproductief werkgedrag. *Gedrag & Organisatie, 27*(4), 407-427. <https://doi.org/10.5117/2014.027.004.407>
- De Vries, R. E., Ashton, M. C., & Lee, K. (2009). De zes belangrijkste persoonlijkheidsdimensies en de HEXACO persoonlijkheidsvragenlijst. *Gedrag & Organisatie, 22*(3), 232-274. <https://doi.org/10.5117/2009.022.003.004>

- De Vries, R. E., Tybur, J. M., Pollet, T. V., & Van Vugt, M. (2016). Evolution, situational affordances, and the HEXACO model of personality. *Evolution and Human Behavior*, 37(5), 407-421. <https://doi.org/10.1016/j.evolhumbehav.2016.04.001>
- De Vries, R. E., Wawoe, K. W., & Holtrop, D. (2015). What is engagement? Proactivity as the missing link in het HEXACO model of personality. *Journal of Personality*, 84(2), 178-193. <https://doi.org/10.1111/jopy.12150>
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness defining "gamification". In *Proceedings of the 15<sup>th</sup> International Academic MindTrek Conference: Envisioning Future Media Environments*, September 28-30, 2011, Tampere, Finland, pp. 9-15. <https://doi.org/10.1145/2181037.2181040>
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41(1), 417-440. <https://doi.org/10.1146/annurev.ps.41.020190.002221>
- Fong, K., & Mar, R. A. (2015). What does my avatar say about me? Inferring personality from avatars. *Personality and Social Psychology Bulletin*, 41(2), 237-249. <https://doi.org/10.1177/0146167214562761>
- Foroughi, C. K., Serraino, C., Parasuraman, R., & Boehm-Davis, D. A. (2016). Can we create a measure of fluid intelligence using Puzzle Creator within Portal 2? *Intelligence*, 56, 58-64. <https://doi.org/10.1016/j.intell.2016.02.011>
- Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6(3), 347-369. <https://doi.org/10.1006/game.1994.1021>
- Forsyth, D. R. (1980). A taxonomy of ethical ideologies. *Journal of Personality and Social Psychology*, 39(1), 175-184. <https://doi.org/10.1037/0022-3514.39.1.175>
- Georgiou, K., & Nikolaou, I. (2020). Are applicants in favor of traditional or gamified assessment methods? Exploring applicant reactions towards a gamified selection method. *Computers in Human Behavior*, 109, 106356. <https://doi.org/10.1016/j.chb.2020.106356>
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48(1), 26-34. <https://doi.org/10.1037/0003-066X.48.1.26>
- Goldberg, L. R., Johnson, J. A., Ebert, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1), 84-96. <https://doi.org/10.1016/j.jrps.2005.08.007>
- Gorsuch, R. L. (1997). New procedure for extension analysis in exploratory factor analysis. *Educational and Psychological Measurement*, 57(5), 725-740. <https://doi.org/10.1177/0013164497057005001>
- Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, 82(3), 379-398. <https://doi.org/10.1037/0022-3514.82.3.379>

- Güth, W., Schmittberger, R., & Schwartz, R. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4), 367-388. [https://doi.org/10.1016/0167-2681\(82\)90011-7](https://doi.org/10.1016/0167-2681(82)90011-7)
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57(3), 639-683. <https://doi.org/10.1111/j.1744-6570.2004.00003.x>
- Hendriks, A. (2012). *SoPHIE – Software platform for human interaction experiments*. University of Osnabrueck, Working Paper.
- Hilbig, B. E., & Zettler, I. (2009). Pillars of cooperation: Honesty-humility, social value orientation and economic behavior. *Journal of Research in Personality*, 43(3), 516-519. <https://doi.org/10.1016/j.jrp.2009.01.003>
- Hogan, R., & Kaiser, R. B. (2005). What do we know about leadership? *Review of General Psychology*, 9(2), 169-180. <https://doi.org/10.1037/1089-2680.9.2.169>
- Holtrop, D., Oostrom, J. K., Dunlop, P. D., & Runneboom, C. (2021). Predictors of faking behavior on personality inventories in selection: Do the ability and motivation to fake predict faking? *International Journal of Selection and Assessment*, 29(2), 185-202. <https://doi.org/10.1111/ijsa.12322>
- Hu, J., & Connelly, B. S. (2021). Faking by actual applicants on personality tests: A meta-analysis of within subject studies. *International Journal of Selection and Assessment*, 29(3-4), 412-426. <https://doi.org/10.1111/ijsa.12338>
- Knoll, M., Meyer, B., Kroemer, N. B., & Schröder-Abé, M. (2015). It takes two to be yourself: An integrated model of authenticity, its measurement, and its relationship to work-related variables. *Journal of Individual Differences*, 36(1), 38-53. <https://doi.org/10.1027/1614-0001/a000153>
- Laljee, M., Laham, S. M., & Tam, T. (2007). Unconditional respect for persons: A social psychological analysis. *Gruppe. Interaktion. Organisation Zeitschrift für Angewandte Organisationspsychologie (GIO)*, 38(4), 451-464.
- Landers, R. N., Auer, E. M., & Abraham, J. D. (2020). Gamifying a situational judgment test with immersion and control game elements. *Journal of Managerial Psychology*, 35(4), 225-239. <https://doi.org/10.1108/JMP-10-2018-0446>
- Langer, M., König, C. J., & Papatthanasious, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, 27(3), 217-234. <https://doi.org/10.1111/ijsa.12246>
- Lee, K., & Ashton, M. C. (2006). Further assessment of the HEXACO personality inventory: Two new facet scales and an observer report form. *Psychological Assessment*, 18(2), 182-191. <https://doi.org/10.1037/1040-3590.18.2.182>
- Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment*, 25(5), 543-556. <https://doi.org/10.1177/1073191116659134>
- Lee, K., Gizzarone, M., Ashton, M. C. (2003). Personality and the likelihood to sexually harass. *Sex Roles*, 49(1/2), 59-69. <https://doi.org/10.1023/A:1023961603479>



- Marsh, H. W., & Richards, G. E. (1988). The Tennessee self-concept scale: Reliability, internal structure, and construct validity. *Journal of Personality and Social Psychology*, 55(4), 612-624. <https://doi.org/10.1037/0022-3514.55.4.612>
- Marshall, M. A., & Brown, J. D. (2006). Trait aggressiveness and situational provocation: A test of the traits as situational sensitivities (TASS) model. *Personality and Social Psychology Bulletin*, 32(8), 1100-1113. <https://doi.org/10.1177/0146167206288488>
- Monga, M. (2017). "Doing the right thing" in the banking sector: Integrity from an upper echelons perspective. In M. Orlitzky & M. Monds (Eds.), *Integrity in business and management* (pp. 105-135). Routledge.
- Moorman, R. H., Darnold, T. C., & Priesmuth, M. (2013). Perceived leader integrity: Supporting the construct validity and utility of a multi-dimensional measure in two samples. *The Leadership Quarterly*, 24(3), 427-444. <https://doi.org/10.1016/j.leaqua.2013.02.003>
- Moorman, R. H., Darnold, T. C., Priesmuth, M., & Dunn, C. P. (2012). Toward the measurement of perceived leader integrity: Introducing a multidimensional approach. *Journal of Change Management*, 12(4), 383-398. <https://doi.org/10.1080/14697017.2012.728746>
- Moshagen, M., & Hilbig, B. E. (2017). The statistical analysis of cheating paradigms. *Behavioral Research Methods*, 49, 724-732. <https://doi.org/10.3758/s13428-016-0729-x>
- Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin*, 35(12), 1661-1671. <https://doi.org/10.1177/0146167209346309>
- Northouse, P. G. (2013). *Leadership: Theory and practice* (sixth edition). Sage Publications.
- Oostrom, J. K., De Vries, R. E., & De Wit, M. (2019). Development and validation of a HEXACO situational judgment test. *Human Performance*, 32(1), 1-29. <https://doi.org/10.1080/08959285.2018.1539856>
- Padilla, A., Hogan, R., & Kaiser, R. B. (2007). The toxic triangle: Destructive leaders, susceptible followers, and conducive environments. *The Leadership Quarterly*, 18(3), 176-194. <https://doi.org/10.1016/j.leaqua.2007.03.001>
- Palanski, M. E., & Yammarino, F. J. (2007). Integrity and leadership: Clearing the conceptual confusion. *European Management Journal*, 25(3), 171-184. <https://doi.org/10.1016/j.emj.2007.04.006>
- Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality*, 17(4), 251-283. <https://doi.org/10.1002/per.474>
- Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues*. American Psychological Association and Oxford University Press.
- Pletzer, J.-L., Bentvelzen, M., Oostrom, J. K., & De Vries, R. E. (2019). A meta-analysis of the relations between personality and workplace deviance: Big Five versus

- HEXACO. *Journal of Vocational Behavior*, 112, 369-383. <https://doi.org/10.1016/j.jvb.2019.04.004>
- Pommier, E. A. (2010). *The Compassion Scale* (Doctoral dissertation).
- Quiroga, M. Á., Román, F. J., De La Fuente, J., Privado, J., & Colom, R. (2016). The measurement of intelligence in the XXI century using video games. *The Spanish Journal of Psychology*, 19, e89. <https://doi.org/10.1017/sjp.2016.84>
- Reynolds, S. J. (2008). Moral attentiveness: Who pays attention to the moral aspects of life? *Journal of Applied Psychology*, 93(5), 1027-1041. <https://doi.org/10.1037/0021-9010.93.5.1027>
- 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(2), 155-180. <https://doi.org/10.1016/j.hrmr.2006.03.004>
- Saucier, G. (2009). Recurrent personality dimensions in inclusive lexical studies: Indications for a Big Six structure. *Journal of Personality*, 77(5), 1577-1614. <https://doi.org/10.1111/j.1467-6494.2009.00593.x>
- Scherkoske, G. (2013). Whither integrity I: Recent faces of integrity. *Philosophy Compass*, 8(1), 28-39. <https://doi.org/10.1111/j.1747-9991.2011.00475.x>
- Schlenker, B. R. (2008). Integrity and character: Implications of principled and expedient ideologies. *Journal of Social and Clinical Psychology*, 27(10), 1078-1125. <https://doi.org/10.1521/jscp.2008.27.10.1078>
- Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin*, 140(2), 374-408. <https://doi.org/10.1037/a0034418>
- Shoss, M. K., & Strube, M. J. (2011). How do you fake a personality test? An investigation of cognitive models in impression-managed responding. *Organizational Behavior and Human Decision Processes*, 116(1), 163-171. <https://doi.org/10.1016/j.obhdp.2011.05.003>
- Simons, T., Friedman, R., Liu, L. A., & McLean Parks, J. (2007). Racial differences in sensitivity to behavioral integrity: Attitudinal consequences, in-group effects, and "trickle down" among black and non-black employees. *Journal of Applied Psychology*, 92(3), 650-665. <https://doi.org/10.1037/0021-9010.92.3.650>
- Suleiman, R. (1996). Expectations and fairness in a modified ultimatum game. *Journal of Economic Psychology*, 17(5), 531-554. [https://doi.org/10.1016/S0167-4870\(96\)00029-3](https://doi.org/10.1016/S0167-4870(96)00029-3)
- Sweetser, P., & Wyeth, P. (2005). GameFlow: A model for evaluating player enjoyment in games. *ACM Computers in Entertainment*, 3(3), 1-24. <https://doi.org/10.1145/1077246.1077253>
- Tabacchi, M. E., Caci, B., Cardaci, M., & Perticone, V. (2017). Early usage of Pokémon Go and its personality correlates. *Computers in Human Behavior*, 72, 163-169. <https://doi.org/10.1016/j.chb.2017.02.047>

- Tangey, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72(2), 271-324. <https://doi.org/10.1111/j.0022-3506.2004.00263.x>
- Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, 146(1), 30-90. <https://doi.org/10.1037/bul0000217>
- Vazire, S., Naumann, L. P., Rentfrow, P. J., & Gosling, S. D. (2008). Portrait of a narcissist: Manifestations of narcissism in physical appearance. *Journal of Research in Personality*, 42(6), 1439-1447. <https://doi.org/10.1016/j.jrp.2008.06.007>
- Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior*, 29(6), 2568-2572. <https://doi.org/10.1016/j.chb.2013.06.033>
- Walumbwa, F. O., Avolio, B. J., Gardner, W. L., Wernsing, T. S., & Peterson, S. J. (2008). Authentic leadership: Development and validation of a theory based measure. *Journal of Management*, 34(1), 89-126. <https://doi.org/10.1177/0149206307308913>
- Wong, S. S., Lim, S. W. H., & Quinlan, K. M. (2016). Integrity in and beyond contemporary higher education: What does it mean to university students? *Frontiers in Psychology*, 7, 1094. <https://doi.org/10.3389/fpsyg.2016.01094>
- Woodward, C., & Pury, C. L. S. (2007). The construct of courage: Categorization and measurement. *Consulting Psychology Journal: Practice and Research*, 59(2), 135-147. <https://doi.org/10.1037/1065-9293.59.2.135>
- Worth, N. C., & Book, A. S. (2014). Personality and behavior in a massively multiplayer online-roleplaying game. *Computers in Human Behavior*, 38, 322-330. <https://doi.org/10.1016/j.chb.2014.06.009>
- Worth, N. C., & Book, A. S. (2015). Dimensions of video game behavior and their relationships with personality. *Computers in Human Behavior*, 50, 132-140. <https://doi.org/10.1016/j.chb.2015.03.056>
- Zeigler-Hill, V., & Monica, S. (2015). The HEXACO model of personality and video game preferences. *Entertainment Computing*, 11, 21-26. <https://doi.org/10.1016/j.entcom.2015.08.001>
- Zhao, K., & Smillie, L. D. (2015). The role of interpersonal traits in social decision making: Exploring sources of behavioral heterogeneity in economic games. *Personality and Social Psychology Review*, 19(3), 277-302. <https://doi.org/10.1177/1088868314553709>

## Correspondentieadres

Ard J. Barends, Universiteit Leiden, Faculteit Rechtsgeleerdheid, Instituut voor Strafrecht & Criminologie, Postbus 9520, 2300 RA Leiden. [a.j.barends@law.leidenuniv.nl](mailto:a.j.barends@law.leidenuniv.nl)