



The Normative Judgment Test of Honesty-Humility: An Implicit Instrument for Organizational Contexts

Dan Asfar, Janneke K. Oostrom, Mark Van Vugt & Marise Born

To cite this article: Dan Asfar, Janneke K. Oostrom, Mark Van Vugt & Marise Born (2024) The Normative Judgment Test of Honesty-Humility: An Implicit Instrument for Organizational Contexts, Human Performance, 37:1-2, 2-33, DOI: [10.1080/08959285.2023.2291208](https://doi.org/10.1080/08959285.2023.2291208)

To link to this article: <https://doi.org/10.1080/08959285.2023.2291208>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 15 Dec 2023.



[Submit your article to this journal](#)



Article views: 1926



[View related articles](#)



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)



OPEN ACCESS



The Normative Judgment Test of Honesty-Humility: An Implicit Instrument for Organizational Contexts

Dan Asfar^a, Janneke K. Oostrom^b, Mark Van Vugt^a, and Marise Born^{a,c}

^aVrije Universiteit Amsterdam; ^bTilburg University; ^cNorth-West University South Africa, South Africa

ABSTRACT

As traits, motives, and attitudes may partly operate outside of individuals' awareness, implicit instruments hold great promise in organizational contexts. One understudied implicit paradigm is the partially-structured attitude measure (PSAM), which assesses individuals' attributes through their judgments of hypothetical persons described in vignettes. Based on this paradigm, we developed the 18-item Normative Judgment Test to assess the personality trait of Honesty-Humility (the NJT-HH). In four studies, we examined the construct- and criterion-related validity of NJT-HH scores. Across studies, NJT-HH scores were positively related to Honesty-Humility, and not related to the other five HEXACO traits (apart from small exceptions). Scores on the NJT-HH were also positively related to scores on a PSAM of honesty, but not related to scores on PSAMs of dissimilar constructs (Study 2). Furthermore, scores on the NJT-HH were negatively related to counterproductive work behavior and positively related to organizational citizenship behavior and task performance, as measured through self-ratings (Study 3) and supervisor ratings (Study 4). Scores on the NJT-HH also explained unique variance in these outcomes above and beyond Honesty-Humility and the other five HEXACO traits. Altogether, these findings provide initial evidence for the practical value of the NJT-HH in organizational contexts.

KEYWORDS

NJT-HH; Honesty-Humility; personality measure; implicit; validity; work outcomes

There is a growing understanding that the greatest asset of any organization is its employees. Indeed, employee behaviors – specifically, behaviors that are aligned with organizational interests, such as low counterproductive work behavior (CWB) and high organizational citizenship behavior (OCB) – have emerged as a key factor contributing to organizational performance (Bolino et al., 2012; Camara & Schneider, 1994; Vardi & Weitz, 2004). Personality traits are important predictors of these specific employee behaviors (e.g., Barrick & Mount, 1991; Lee et al., 2019; Pletzer et al., 2019, 2021). Hence, to gain insights into which employees will show behaviors that are aligned with organizational interests, that is, who act ethically and cooperatively, numerous organizations worldwide have included personality measures in their assessment programs (Rothstein & Goffin, 2006). In recent years, there has been growing evidence in personality literature that the trait Honesty-Humility (HH) within the HEXACO personality model is one of the strongest and most consistent predictors of CWB and OCB (Lee et al., 2019; Pletzer et al., 2019, 2020, 2021). Assessing trait HH could therefore be beneficial for organizations.

While personality traits are most typically assessed with self-report measures, scholars have called for more research on “innovative techniques that go beyond, without replacing, self-report measures . . .” (Funder, 2002, p. 639; see also Sackett et al., 2017). Among these innovative techniques are

CONTACT Janneke K. Oostrom  J.K.Oostrom@tilburguniversity.edu  Tilburg School of Social and Behavioral Sciences, Department of Social Psychology, Tilburg University, Netherlands

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/08959285.2023.2291208>

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

implicit instruments, which assess traits, motives, and attitudes that people might not be willing to disclose or of which they are unaware (Moors et al., 2010). Implicit instruments have been found to predict relevant work outcomes and explain variance in these outcomes above and beyond the variance explained by self-report measures of the same construct (see Uhlmann et al., 2012). The goal of the present research is to develop an implicit instrument of HH that can be used in organizational contexts. This instrument, which we label the Normative Judgment Test of Honesty-Humility (the NJT-HH), is based on the understudied partially-structured attitude measure (PSAM; Vargas et al., 2004). The PSAM assesses individuals' trait levels through their judgments of ambiguous behaviors of hypothetical persons described in vignettes. Uhlmann et al. (2012) suggested that this paradigm is "useful for predicting who is likely to engage in high levels of organizational citizenship or volunteering . . . or for discerning an individual's standards for ethical behavior or corporate social responsibility" (p. 579). To address this suggestion, we conducted four studies to investigate the construct- and criterion-related validity of NJT-HH scores for predicting CWB and OCB. In Study 4, we also exploratorily examine the relationship between NJT-HH scores and task performance. We demonstrate that the NJT-HH is a promising complement or alternative to self-report measures for predicting CWB, OCB, and task performance.

The importance of Honesty-Humility in organizational contexts

To be able to prevent harmful and promote desirable workplace behaviors, much research has been devoted to the predictors of CWB and OCB (Berry et al., 2007; Harari & Viswesvaran, 2018). CWB has been defined as "voluntary behavior that violates significant organizational norms and in so doing threatens the well-being of an organization, its members, or both" (Robinson & Bennett, 1995, p. 556) and OCB is defined as "individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and that in the aggregate promotes the effective functioning of the organization" (Organ, 1988, p. 4). Several organizational characteristics have been associated with employees' CWB and OCB, including organizational justice, leadership style, and team empowerment (Colquitt et al., 2013; Kirkman & Shapiro, 2001; Mitchell & Ambrose, 2007). Furthermore, individual differences as predictors of CWB and OCB, particularly those in personality, have received increased attention in recent years (Lee et al., 2019; Pletzer et al., 2020, 2021).

Personality is most commonly described in terms of the Big Five (or FFM) dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism/Emotional Stability. However, in the last two decades, an increasing number of studies in the personality literature have indicated that personality might be more optimally described in terms of six, instead of five, dimensions (Ashton et al., 2004; De Raad et al., 2014; Saucier, 2009). The most studied six-dimensional personality framework is the HEXACO model, which consists of the dimensions Honesty-Humility (HH), Emotionality (E), eXtraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O) (Ashton & Lee, 2007; Ashton et al., 2014). While the HEXACO and the Big-Five model differ somewhat in the conceptualization of the dimensions of Agreeableness and Emotional Stability/Emotionality, the main difference between the personality models is the addition of the Honesty-Humility (HH) dimension in the HEXACO (e.g., see Ashton & Lee, 2020).

HH is defined as "the tendency to be fair and genuine in dealing with others, in the sense of cooperating with others even when one might exploit them without suffering retaliation" (Ashton & Lee, 2007, p. 156). This trait captures an individual's tendency to refrain from manipulation, fraud, and exploitation (Ashton & Lee, 2007; Lee & Ashton, 2004). In line with HH's description, this trait correlates positively with overt and personality-based integrity measures, with typical correlations exceeding $r = .50$ (e.g., Lee et al., 2005, 2019; Marcus et al., 2007). The propensity of individuals who score low on HH to deceive and exploit others makes them also "more likely to behave in their own interest at the expense of the best interest of their employer" (Oh et al., 2011, p. 500). Indeed, recent meta-analytic work has shown that HH is a good predictor of CWB ($r = -.35$ to $-.39$) and explains unique variance in CWB, above and beyond the variance explained by the other five HEXACO traits

(Lee et al., 2019; Marcus et al., 2007; Pletzer et al., 2019, 2020). Furthermore, HH is positively related to ethical leadership (De Vries, 2012), and negatively to delinquent work behaviors (De Vries & Van Gelder, 2015; Lee et al., 2005) and unethical business decisions (Ashton & Lee, 2008; De Vries et al., 2017). Individuals who score high on HH also tend to feel responsible for behaving prosocially toward others (Hilbig & Zettler, 2009; Oh et al., 2014). In organizational contexts, high HH employees have a proclivity to behave cooperatively toward colleagues and to adhere to organizational rules (Ashton & Lee, 2007; Bourdage et al., 2012), because they believe that it is their moral responsibility to do so (Marcus et al., 2007). Correspondingly, meta-analyses have revealed that HH is a positive predictor of OCB ($r = .10$ to $.18$; Lee et al., 2019; Pletzer et al., 2021).

Implicit instruments

A considerable amount of empirical work has shown that people process information about themselves and their environment not only explicitly (i.e., controlled or conscious), but also implicitly (i.e., automatic or unconscious; Bargh & Chartrand, 1999; Dijksterhuis & Nordgren, 2006; Epstein, 1994; Fazio, 1990; Greenwald & Banaji, 1995). To assess such unconscious psychological attributes, several implicit instruments have been developed and empirically investigated (for an overview, see Uhlmann et al., 2012). Common implicit instruments are the picture story exercise (PSE; Schultheiss et al., 2008; similar to the thematic apperception test [TAT]; Morgan & Murray, 1935), the implicit association test (IAT; Greenwald et al., 1998), and the conditional reasoning test (CRT; James, 1998). There is substantial empirical evidence that scores on these instruments predict employee behaviors and explain unique variance in these behaviors above and beyond the variance explained by scores on self-report measures (Apers et al., 2019; Dietl & Meurs, 2019; Galić et al., 2014; Lang et al., 2012; Leavitt et al., 2011). Furthermore, due to their indirect nature, these implicit instruments have been found to be resistant to response distortion (faking) under specific conditions (e.g., LeBreton et al., 2007; Steffens, 2004; Vecchione et al., 2014; Wiita et al., 2020).

Despite the positive findings regarding the criterion-related validity and fakability of scores on these commonly studied implicit instruments, they also have some important limitations. The PSE (or TAT) is time-consuming to administer and score and it requires extensive training to score the subjective interpretation of participants' responses. Furthermore, the PSE (or TAT) lacks face validity, leading to defensive responses by applicants, and its reliability and validity remain debated (Lilienfeld et al., 2000). IAT scores have poor test-retest reliability (Cunningham et al., 2001; Egloff et al., 2005; LeBel & Paunonen, 2011), and participants have trouble seeing the IAT's job relevance and feel like they have little opportunity to perform on this test (Wright & Meade, 2011). Additionally, in one recent study, scores on an IAT of HH showed no criterion-related validity (Van Rensburg et al., 2022). Finally, the CRT for aggression has a highly skewed distribution of test scores with a mean of 3.89 on a scale of 0 (no aggression) to 22 (extremely aggressive) (James & LeBreton, 2012), which makes it difficult for this test to discriminate amongst individuals with low levels of aggression (DeSimone & James, 2015).

The issues with these commonly studied implicit instruments encouraged us to address the call for the development of novel implicit psychological instruments (Funder, 2002; Sackett et al., 2017). As we will explain and show in the remainder of this manuscript, the implicit instrument that we introduce addresses most of the aforementioned issues: The NJT-HH is relatively easy to administer, is automatically scored and standardized, and has no skewed distribution of test scores. Importantly, we demonstrate that NJT-HH scores can predict employee behaviors.

The partially-structured attitude measure

In 2004, Vargas et al. introduced a new paradigm of implicit testing: the partially-structured attitude measure (PSAM). In the PSAM paradigm, individuals judge trait levels of hypothetical persons who are described in vignettes. This paradigm is based on the phenomenon that the self serves as an anchor

in social judgment (cf. Sherif & Hovland, 1961; for an overview, see; Dunning, 2012). Hence, individuals' judgments of the persons in the vignettes are an indication of their own trait levels. Research has shown that people judge others in ways that promote their self-esteem (Beauregard & Dunning, 1998), such that people who are high on a desirable attribute are more judgmental of people who are low on the desirable attribute (Beauregard & Dunning, 1998; Dunning & Cohen, 1992; Dunning & Hayes, 1996; Eidelman & Biernat, 2007; Protzko & Schooler, 2019). For example, Dunning and Cohen (1992) measured students' athleticism, math ability, punctuality, studiousness, and how well-read they are, and asked them to judge the levels of several hypothetical persons on these characteristics. The authors found that students with higher scores on these characteristics judged the hypothetical persons lower on the corresponding characteristics (except for punctuality). Correspondingly, the series of studies by Vargas et al. (2004) demonstrated that trait level estimations of hypothetical persons in vignettes that pertain to honesty, political orientation, and religiosity were inversely associated with participants' self-reported and actual behaviors relevant to these domains. For example, participants who perceived hypothetical persons who engaged in ambiguous dishonest behavior as very dishonest, were themselves less likely to cheat on anagrams. Importantly, these judgments explained unique variance in corresponding outcomes above and beyond scores on explicit measures of the same constructs.

In the present research, we develop an implicit instrument of HH, the NJT-HH, that is based on the PSAM paradigm (Vargas et al., 2004). The NJT-HH is different from Vargas et al. (2004) PSAM of honesty, as the NJT-HH assesses HH, which is a broader construct than honesty, and contains a larger number of items than Vargas et al.'s (2004) measure, increasing the reliability of its scores.

The present research

In the current research, we investigated the construct- and criterion-related validity of scores on the NJT-HH. The construct-related validity of NJT-HH scores is examined through their relationship with the HEXACO traits (Lee & Ashton, 2004). Scores on implicit instruments generally show modest positive correlations with scores on explicit measures of the same construct (e.g., Bosson et al., 2000; Hofmann et al., 2005; Vargas et al., 2004). These modest correlations between implicit and explicit instruments are likely due to motivational biases in the report of consciously accessible representations (Fazio & Olson, 2003), a lack of introspective access to implicitly assessed representations (Greenwald & Banaji, 1995), and method-related characteristics of the two instruments (e.g., type of responses; see Payne et al., 2008). We posit the following hypothesis for the convergent validity of NJT-HH scores:

Hypothesis 1 (H1): NJT-HH scores are modestly and positively correlated with HEXACO HH scores.¹

The NJT-HH was developed to exclusively measure trait HH and no other personality traits. Thus, we posit the following hypothesis regarding the discriminant validity of scores on the NJT-HH:

Hypothesis 2 (H2): NJT-HH scores are not significantly correlated with scores on any of the five other HEXACO traits.

The construct-related validity of NJT-HH scores is also examined through their relationship with scores on the three PSAMs developed by Vargas et al. (2004), which aim to measure honesty, political conservatism, and religiosity. Although convergent correlations between scores on different implicit measures are generally weak (Bosson et al., 2000), they tend to be higher when the underlying constructs and structural features of the measure are more alike (Rudolph et al., 2008). Thus, although the NJT-HH measures HH, which is a broader construct than honesty in the PSAM, we expect a modest correlation between scores on the

¹Following Cohen (1988), we regard $r = .10$, $r = .30$, and $r = .50$, respectively as small, moderate, and large correlations.

two measures as they target partly overlapping constructs and should reflect similar implicit processes. Hence, we posit the following hypothesis for the convergent validity of NJT-HH scores:

Hypothesis 3 (H3): NJT-HH scores are modestly and positively correlated with scores on the PSAM of honesty.

As the underlying constructs of the NJT-HH and the PSAM for political conservatism and religiosity do not overlap, we do not expect the scores on these measures to be correlated. In fact, if NJT-HH scores would show significant correlations with scores on these measures, this would indicate that NJT and PSAM scores merely measure idiosyncratic response tendencies to vignettes (Hopkins & King, 2010; King & Wand, 2007) rather than substantive constructs. As there is ample evidence that people reveal their own traits when judging others (e.g., Beauregard & Dunning, 1998; Dunning & Hayes, 1996; Oostrom et al., 2017; Protzko & Schooler, 2019), we posit that NJT-HH scores indicate trait levels of HH rather than idiosyncratic response tendencies. Hence, we propose the following hypothesis regarding the discriminant validity of NJT-HH scores:

Hypothesis 4 (H4): NJT-HH scores are not significantly correlated with scores on the PSAMs of political conservatism and religiosity.

People who score high on HH tend to be honest, sincere, and greed avoidant. These are integrity-related traits and should be associated with the avoidance of undesirable and harmful behaviors. Indeed, HH has been shown to be a strong negative predictor of CWB (Lee et al., 2019; Marcus et al., 2007; Pletzer et al., 2019, 2020). Furthermore, Vargas et al. (2004) showed that the PSAM of honesty is negatively related to cheating on anagrams. Hence, for the criterion-related validity of NJT-HH scores, we posit the following hypothesis:

Hypothesis 5 (H5): NJT-HH scores are negatively correlated with CWB.

The personality trait of HH has also been conceptually linked with OCB (Lee et al., 2019). People who score high on HH tend to have a high moral conscience and feel responsible for behaving prosocially toward others (Hilbig & Zettler, 2009; Oh et al., 2014). Correspondingly, meta-analyses have shown that HH is a positive predictor of OCB (Lee et al., 2019; Pletzer et al., 2021). Hence, we posit the following hypothesis:

Hypothesis 6 (H6): NJT-HH scores are positively correlated with OCB.

As explained earlier, scores on implicit instruments capture unique variance in attributes to scores on explicit self-report measures of the same construct (e.g., Hofmann et al., 2005). Hence, scores on implicit instruments can explain variance in work outcomes above and beyond the variance explained by scores on traditional self-report measures (for an overview, see Uhlmann et al., 2012). In line with these findings, we propose the following two hypotheses:

Hypothesis 7 (H7): NJT-HH scores explain unique variance in CWB above and beyond the variance explained by HEXACO HH scores.

Hypothesis 8 (H8): NJT-HH scores explain unique variance in OCB above and beyond the variance explained by HEXACO HH scores.

Study 1

We developed the NJT-HH following test development recommendations by Lane et al. (2016). First, we generated items to measure trait HH based on the definition of this trait within the HEXACO model (Ashton & Lee, 2007; Ashton et al., 2014) and the method of Vargas et al. (2004). Next, we determined the factor structure of NJT-HH scores using both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) and verified their internal consistency. We combined all NJT-HH data from Studies 2–4 and our test-retest study and randomly divided the total dataset into a training sample and a test sample. We reduced our items based on the results of the factor and internal consistency analyses in our training sample, and cross-validated these results in our test sample. Last, we reported the test-retest reliability of scores on the final NJT-HH.

Method study 1

The development of the NJT-HH and the construction of its items was an iterative process. Together with three master's students, the first author of this manuscript designed 23 items in total. The items have the form of vignettes, followed by a question about the hypothetical person described in the vignette (for example items, see Table 1).² As Dunning and Cohen (1992) demonstrated that participants varied mostly in their evaluation of hypothetical persons who score low rather than high on an attribute, all vignettes described a hypothetical person who engaged in ambiguous behavior that can be considered low on HH.

Table 1. Example items of the normative judgment test of honesty-humility (NJT-HH).

Jane works in a lunchroom. In front of the lunchroom, there are usually two product promotors, distributing products like chocolate and drinks. Jane gives them a free lunch, and in return, they give her many of the products that they are supposed to distribute to the people outside. Jane takes these products home with her. How dishonest do you consider Jane to be? 1 = not at all dishonest 2 = a little dishonest 3 = moderately dishonest 4 = very dishonest 5 = completely dishonest
Thom is the owner of a repair shop. He is the only person in town who can fix battery problems with electric cars. When customers ask him to fix a battery problem with their car, Thom requests three times the amount of money than he would otherwise ask for a similar repair job. How greedy do you consider Thom to be? 1 = not at all greedy 2 = a little greedy 3 = moderately greedy 4 = very greedy 5 = completely greedy
An amateur soccer team wins a tournament and is awarded a big trophy. There are no rules about who should keep the trophy. Andrew thinks he is the best player in the team and he therefore asks the coach to keep the trophy for himself. How self-centered do you consider Andrew to be? 1 = not at all self-centered 2 = a little self-centered 3 = moderately self-centered 4 = very self-centered 5 = completely self-centered

²For proprietary reasons, we are unable to provide the full list of items. Researchers who want to use the NJT-HH can contact the test publisher by sending a request to info@noa-vu.nl.

The items pertain to situations at work (11 items), but also other life domains (6 items). Furthermore, to capture all HH aspects, each item relates to one of the four facets of HH (Sincerity, Fairness, Modesty, and Greed avoidance). However, as our goal was to develop a relatively short instrument, we intended to measure the overall construct of HH rather than its four underlying facets. Next, the authors and students had a group discussion about the clarity, face validity, and potential risk of bias of each item. This resulted in relatively small adjustment of most of the items.

Then, to assess the content-related validity (Colquitt et al., 2019), the item pool was reviewed on these same aspects by fourteen I/O psychologists who work in a consultancy firm that develops psychological tests and assessments. The I/O psychologists had 2 to 23 years ($M = 9.0$, $SD = 7.73$) of work experience in the field of test development. Specifically, through a survey, we asked the reviewers to indicate for each item (a) whether it was clearly formulated and easy to understand in terms of word usage, complexity, and ambiguity (and if not, why this might be the case), (b) to what extent – using a 5-point Likert scale – it measures each of the six HEXACO personality dimensions, (c) if it might function differently for certain groups, such as men and women or different ethnic groups (and if so, why this might be the case), and (d) whether the reviewers had suggestions to improve the item. Based on this content-related validity study we adjusted the items: Formulations of fourteen items were adjusted to make them clearer. Formulations of thirteen items were adjusted because they could potentially measure other personality traits. Finally, five items were revised because these items might have been biased against particular social groups or might have been differently interpreted by different groups. In the Supplementary Materials, an illustration is provided of how an item has been revised.

Next, we focused on the adjectives that we could use in the question that followed each of the vignettes (e.g., How [*adjective*, e.g., dishonest] do you consider Jane to be?). We favored custom adjectives for each item to default adjectives, as the hypothetical person could be more adequately judged when the adjective matches the vignette. To adopt the most applicable adjectives, we first reviewed Ashton and Lee's (2008) list of adjectives that load on HH, and then selected nine adjectives that are clear and easy to comprehend and that are applicable (e.g., dishonest, selfish).³ Next, we presented 31 Amazon Mechanical Turk (MTurk) workers with the NJT-HH items together with the nine adjectives. We asked them to indicate which adjective, according to them, is the most applicable to the hypothetical person in the vignette. For each item, we used the adjective that was most often selected.

In the NJT-HH, participants were asked to judge the hypothetical person's level of Honesty-Humility described in the vignette using a 5-point Likert scale. The item response options varied from 1 = *not at all* [adjective] to 5 = *completely* [adjective]. The adjectives were phrased negatively (e.g., dishonest, immodest). Higher scores were therefore anticipated to indicate higher levels of trait HH.

We combined all NJT-HH data from Studies 2–4 and the first measurement of the NJT-HH in our test-retest study (see below). This resulted in a total dataset of 700 participants ($M_{\text{age}} = 35.94$ years, $SD = 12.78$; 49.9% identified as men, 49.3% as women, and 0.9% as other or did not report their gender⁴). Most participants had obtained a bachelor's degree ($n = 311$; 44.4%), a high school degree ($n = 135$; 19.3%), or a master's degree ($n = 125$; 17.9%). We randomly divided the total dataset into a training sample ($n = 350$) and a test sample ($n = 350$).

To provide evidence for the temporal consistency of NJT-HH scores, we conducted a small online study among 140 Prolific workers who completed the NJT-HH. Two and a half weeks later, we asked the participants to complete the NJT-HH again. In total, 104 participants (53.8% female; $M_{\text{age}} = 35.35$, $SD = 12.88$) completed the NJT-HH twice (i.e., a retention rate of 74.3%). The average time interval between the two test administrations was 15.71 days ($SD = 1.00$, min = 15 days, max = 19 days).

³The list of adjectives consisted of Dishonest, Insincere, Greedy, Arrogant, Self-centered, Selfish, Immodest, Egoistical, and Untruthful. The former four adjectives were most frequently selected and used in the items.

⁴Note that we explicitly asked participants to report their gender and not their sex in all studies.

Table 2. EFA factor loadings, CFA factor loadings, and corrected item-total correlations in training sample and test sample (study 1).

Item	Training sample (<i>n</i> = 350)			Test sample (<i>n</i> = 350)		
	EFA factor loading	CFA factor loading	Corrected item-total correlation	EFA factor loading	CFA factor loading	Corrected item-total correlation
1	.28	.29	.22			
2*	.34	.36	.31	.31	.34	.28
3	.26	.27	.21			
4*	.48	.48	.36	.44	.52	.40
5*	.30	.34	.28 ^a	.36	.37	.34
6	.24	.27	.25			
7*	.47	.52	.41	.45	.51	.40
8*	.53	.61	.50	.47	.50	.40
9*	.46	.51	.38	.33	.30	.30
10*	.55	.60	.42	.60	.66	.54
11*	.46	.53	.41	.60	.66	.52
12*	.60	.63	.52	.57	.57	.50
13*	.53	.60	.47	.60	.62	.54
14*	.37	.32	.33	.44	.41	.38
15*	.36	.33	.35	.33	.11	.30
16*	.50	.52	.46	.58	.58	.52
17	.26	.18	.29			
18*	.42	.39	.33	.48	.53	.45
19*	.47	.39	.42	.48	.46	.42
20*	.39	.38	.36	.28	.27	.27
21*	.51	.50	.47	.36	.35	.34
22*	.42	.39	.40	.36	.30	.33
23	.29	.28	.28			

Note. EFA = Exploratory factor analysis; CFA = confirmatory factor analysis. * Item is included in the final 18-item NJT-HH. The CFA factor loadings are the item loadings on the general factor. ^a Even though this item has a relatively low corrected item-total correlation, we decided to retain this item for two reasons: 1) without this item the total number of items intended to measure sincerity would drop to two items, and 2) the item does have an adequate EFA and CFA factor loading.

Results study 1

Table 2 presents the results of the EFA, CFA, and internal consistency analysis for both the training and test sample.

Training sample

We conducted an EFA (with SPSS version 27) with the 23 NJT-HH-items using the principal axis factoring extraction method (e.g., Chuah et al., 2006). The Kaiser – Meyer–Olkin measure (Kaiser, 1970) verified the sampling adequacy for the analysis, KMO = .79 (see Hutcheson & Sofroniou, 1999). Bartlett's test of sphericity (χ^2 [253] = 1172.72, $p < .001$) indicated that the correlations between items were sufficiently large for the principal axis factoring method. The scree plot showed an inflexion that would justify retaining one factor, which explained 20.92% variance.

To further examine the factor structure of the NJT-HH scores, we conducted a CFA (with Amos version 26.0), using the maximum likelihood estimation method and full information maximum likelihood (FIML) to deal with missing data. To examine the extent to which scores on the NJT-HH items reflect the overall construct of HH or its facets, we used a bi-factor approach (Gustafsson & Åberg-Bengtsson, 2010; Reise et al., 2010). The bi-factor model showed a reasonable fit to the data (χ^2 [207] = 401.01, $p < .001$, CFI = .83, TLI = .77, RMSEA = .05 [90% CI = .04; .06]). The average item loading on the general factor (.42) was substantially higher than the average item loading on the subscales (.15), confirming that the item scores primarily reflect the general construct of HH rather than its facets. Furthermore, the fit of the bi-factor model was significantly better than the fit of a second-order model with HH as the second-order factor and its four facets as the first-order factors (χ^2 [226] = 449.98,

$p < .001$, CFI = .80, TLI = .75, RMSEA = .05 [90% CI = .05; .06]; $\Delta\chi^2[19] = 48.97$, $p < .001$), as well as the fit of a single-order four-factor model ($\chi^2[224] = 437.11$, $p < .001$, CFI = .81, TLI = .76, RMSEA = .05 [90% CI = .04; .06]; $\Delta\chi^2[17] = 36.10$, $p < .001$), with latent correlations varying between $r = .50$ and $r = .77$.

Finally, we conducted an internal consistency analysis and calculated the corrected item-total correlations. Coefficient alpha of the scores on the 23-item NJT-HH was .82. As can be expected, coefficient alphas of the subscale scores were relatively low (average $\alpha = .60$).

We reduced the number of items of the NJT-HH based on the results of the factor and internal consistency analyses. Items with a factor loading below .30 in either the EFA or CFA (on the general factor) or a corrected item-total correlation below .30 were removed. Six items met this criterion. However, we decided to retain one of these items to ensure that each HH facet was covered by at least three items. Hence, the final NJT-HH consisted of 18 items.

Test sample

To cross-validate our results, we repeated the EFA, CFA, and internal consistency analyses for the 18-item NJT-HH in our test sample. The KMO of .82 verified the sampling adequacy for the EFA analysis, and Bartlett's test of sphericity ($\chi^2 [153] = 832.71$, $p < .001$) indicated that the correlations between items were sufficiently large for the principal axis factoring method. As in the training sample, the scree plot showed an inflexion that would justify retaining one factor, which explained 25.22% variance. All factor loadings were above .30, except for one item with a factor loading of .28.

The CFA analyses indicated that the bi-factor model ($\chi^2[117] = 213.71$, $p < .001$, CFI = .89, TLI = .84, RMSEA = .05 [90% CI = .04; .06]) showed a reasonable fit to the data. All factor loadings were above .30, except for two items with a factor loading of .27 and .11. The bi-factor model showed a better fit to the data than the second-order model ($\chi^2[131] = 246.33$, $p < .001$, CFI = .87, TLI = .82, RMSEA = .05 [90% CI = .04; 0.6]); $\Delta\chi^2[14] = 32.62$, $p < .001$) and the single-order four-factor model ($\chi^2[129] = 238.22$, $p < .001$, CFI = .87, TLI = .83, RMSEA = .05 [90% CI = .04; .06]; $\Delta\chi^2[12] = 24.51$, $p = .017$), with latent correlations varying between $r = .59$ and $r = .84$.

Finally, we conducted an internal consistency analysis and calculated the corrected item-total correlations. Coefficient alpha of the scores on the 18-item NJT-HH was .81. All items had a corrected item-total correlation of .30 or higher, except for two items with corrected item-total correlations just below .30 (i.e., .27 and .28). Again, the coefficient alphas of the subscale scores were relatively low (average $\alpha = .56$).

Test-retest reliability

The test-retest reliability of NJT-HH scores was $r = .68$ (and $r = .80$ corrected for unreliability). This test-retest reliability is comparable to the test-retest reliability of scores on the CRT for aggression (James & LeBreton, 2012), and higher than the test-retest reliability of scores on several other implicit instruments, such as the IAT (Cunningham et al., 2001; Egloff et al., 2005; LeBel & Paunonen, 2011) and the TAT (Lilienfeld et al., 2000).

Discussion study 1

We tested the NJT-HH scores' factor structure and internal consistency in a training and test sample. Overall, the test sample validated the results of the training sample by demonstrating adequate factor structure and internal consistency for scores on the 18-item NJT-HH, measuring the overall construct of HH. Furthermore, we found adequate test-retest reliability for scores on the 18-item NJT-HH. As

we did not intend to measure HH facets with the NJT-HH, the item scores primarily reflect the general construct of HH rather than its facets, and the alpha coefficients of subscale scores are relatively low, we only used the overall NJT-HH score in the subsequent studies.

Study 2

The goal of Study 2 was to examine the construct-related validity of NJT-HH scores by examining their relationship with scores on the HEXACO scales and the PSAMs of honesty, political conservatism, and religiosity.

Method study 2

Sample and procedure

We recruited a sample of working adults (18–65 years old) via Prolific, a crowdworking platform dedicated to academic research (Peer et al., 2017). We targeted native English-speaking participants living in the United Kingdom or the United States. A power analysis (Faul et al., 2009) showed that the minimum required sample size was 142 to have 90% power to detect a small- to medium-sized convergent correlation coefficient ($r = .24$; based on Hofmann et al., 2005), with $\alpha = .05$. As we expected some failed attention checks, we oversampled participants to account for possible exclusions. Hence, we ceased recruitment after about 200 participants completed the study.

The study consisted of five measures: the NJT-HH, the HEXACO-60 (Ashton & Lee, 2009), and the three PSAMs of honesty, political conservatism, and religiosity (Vargas et al., 2004). The study was approved by the faculty's ethics committee. In the consent form, participants were informed about their rights and the anonymous data processing and the confidential treatment of their data. At the end of the study, we explained the goal of the study to the participants and thanked them for their participation. The study took about 35 minutes to complete. We paid our participants £5 or \$6.

We included two attention check items (e.g., "Please select strongly agree on this item"). A total of 202 participants completed our study. After removing four participants who failed our attention checks, our final sample consisted of 198 participants ($M_{\text{age}} = 40.70$ years, $SD = 12.80$; 32.8% identified as men, 66.7% as women, and 0.5% as other). Most participants had obtained a bachelor's degree ($n = 77$; 38.9%), a high school degree ($n = 62$; 31.3%), or a master's degree ($n = 30$; 15.2%). The participants had up to 52 years of work experience ($M = 19.22$, $SD = 12.60$), and worked up to 60 hours per week ($M = 28.39$, $SD = 15.84$).

Measures

Normative Judgment Test of Honesty-Humility (NJT-HH)

We used the 18-item NJT-HH ($\alpha = .79$). The KMO of .78 and Bartlett's test of sphericity ($\chi^2 [153] = 688.76$, $p < .001$) indicated that the data were suitable for an EFA analysis (with SPSS version 27). The scree plot indicated one factor, which explained 23.08% variance. Factor loadings varied between .24 and .58.

HEXACO

We used the HEXACO-60 (Ashton & Lee, 2009; De Vries et al., 2009) to measure the six HEXACO traits. Items were rated on a 5-point Likert scale (1 = *strongly disagree*; 5 = *strongly agree*). Example items are "I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed" (H), "I have to cry during sad or romantic movies" (E), "I easily approach strangers" (X), "Even when I'm treated badly, I remain calm" (A), "I plan ahead and organize things, to avoid scrambling at the last minute" (C), and "I like people with strange ideas" (O). The alpha coefficients in the current study were .78 for HH, .81 for E, .84 for X, .80 for A, .75 for C, and .81 for O.

The Partially Structured Attitude Measure (PSAM)

We used the three PSAMs of Vargas et al. (2004) to measure honesty (6 items), political conservatism (6 items), and religiosity (20 items). Each of the PSAMs consists of short vignettes describing different individuals engaging in ambiguous behaviors related to the construct of interest. Following the honesty vignettes, participants responded to three questions: “How dishonest was the behavior [name of individual] performed?,” “How dishonest do you think [name of individual] is, in general?,” and “Out of 100 people, how many would do what [name of individual] did in that situation?.” Participants responded to the first two questions using an 11-point Likert scale (1 = *not at all dishonest*; 11 = *extremely dishonest*), and to the third question by providing a number from 0 to 100. Responses to the third question were reverse scored, and responses to all three questions for each of the vignettes were transformed to *z*-scores and then averaged to create an overall score, with higher scores indicating higher honesty. Following the political conservatism vignettes, participants responded to two questions, “How politically conservative/liberal was the behavior [name of individual] performed?” and “How politically conservative/liberal do you think [name of individual] is, in general?,” using an 11-point Likert scale (1 = *very liberal*; 11 = *very conservative*). Responses were reverse scored and then combined into a single score, with higher scores indicating higher conservatism. Following the religiosity vignettes, participants also responded to two questions, “How religious was the behavior [name of individual] performed?” and “How religious do you think [name of individual] is, in general?,” using an 11-point Likert scale (1 = *not at all religious*; 11 = *very religious*). Responses were reverse scored and then combined into a single score, with higher scores indicating higher religiosity. The alpha coefficients in the current study were .61 for honesty, .45 for political conservatism, and .85 for religiosity. The low internal consistency of the scores on political conservatism is most likely due to the vignettes describing political debates or events that were relevant 20 years ago, but that people are no longer familiar with (e.g., the Whitewater scandal). Hence, the correlations with the scores on the PSAM of political conservatism need to be interpreted with caution.

Results study 2

Means, standard deviations, and bivariate intercorrelations of the sociodemographic variables, the NJT-HH, the HEXACO scales, and the PSAMs are presented in Table 3. We first compared the correlations of the scores on the NJT-HH, HEXACO HH, and PSAM-H with age and gender. The correlation between HEXACO HH scores and age ($r = .31, p < .001$) was significantly stronger than the

Table 3. Means, standard deviations, and bivariate intercorrelations of variables in study 2.

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Gender	0.67	0.47	(-)												
2. Age	40.70	12.80	-.01	(-)											
3. WH	28.39	15.84	-.10	-.18*	(-)										
4. NJT-HH	3.37	0.49	.13	.04	-.08	(.79)									
5. HH	3.51	0.67	.10	.31**	-.13	.25**	(.78)								
6. E	3.40	0.68	.40**	-.08	-.11	.10	-.03	(.81)							
7. X	2.97	0.72	-.12	.02	.05	-.06	-.08	-.33**	(.84)						
8. A	3.17	0.65	-.11	.19**	-.07	.16*	.41**	-.12	.16*	(.80)					
9. C	3.66	0.54	.04	.13	.00	.03	.28**	-.10	.17*	.14	(.75)				
10. O	3.52	0.69	-.06	.19**	-.09	-.04	.15*	-.07	.11	.19**	.15*	(.81)			
11. PSAM-H	0.00	0.47	.18*	.01	.03	.40**	.23**	.09	.05	.15*	.02	.14*	(.61)		
12. PSAM-PC	5.61	0.93	-.21**	-.09	-.03	-.05	-.05	-.15*	.00	.04	.03	-.10	-.26**	(.45)	
13. PSAM-R	6.22	1.09	-.03	.17*	-.07	-.07	.07	-.05	.14*	-.04	.05	-.15*	-.10	.05	(.85)

Note. Gender: male = 0, female = 1. WH = Working hours, NJT-HH = Normative Judgment Test of Honesty-Humility, HH = Honesty-Humility, E = Emotionality, X = Extraversion, A = Agreeableness, C = Conscientiousness, O = Openness to Experience, PSAM-H = Partially Structured Attitude Measure of Honesty, PSAM-PC = Partially Structured Attitude Measure of Political Conservatism, and PSAM-R = Partially Structured Attitude Measure of Religiosity. The scores for the NJT-HH and HEXACO traits range from 1 to 5. The PSAM-H scores represent averaged *z*-scores. The scores on the PSAM-PC and PSAM-R range from 1 to 11. *N* = 198.

* $p < .05$; ** $p < .01$ (two-tailed).

correlation between the NJT-HH scores and age ($r = .04, p = .535; z = 3.17, p < .001$) and the PSAM-H scores and age ($r = .01, p = .901; z = 3.49, p < .001$). We found significant gender differences on the PSAM-H scores, favoring women ($M = -0.12, SD = 0.45$ for men; $M = 0.06, SD = 0.47$ for women; $t[195] = 2.50, p = .013, d = 0.38$), but no gender differences on the NJT-HH scores ($M = 3.28, SD = 0.45$ for men; $M = 3.42, SD = 0.50$ for women; $t[195] = 1.88, p = .062, d = 0.28$) and HEXACO HH scores ($M = 3.41, SD = 0.71$ for men; $M = 3.56, SD = 0.65$ for women; $t[195] = 1.45, p = .148, d = 0.22$).

Hypothesis testing

We hypothesized that NJT-HH scores correlate modestly and positively with HEXACO HH scores (H1), but not with the scores on the other five HEXACO scales (H2). The results revealed a significant modest and positive correlation between the NJT-HH and HEXACO HH scores ($r = .25, p < .001$), supporting H1. We also found a modest and positive correlation between the NJT-HH and HEXACO A scores ($r = .16, p = .021$). However, NJT-HH scores were not significantly correlated with any of the scores on the other HEXACO scales, largely supporting H2. We also hypothesized that NJT-HH scores correlate positively with scores on the PSAM of honesty (H3), but not with scores on the PSAMs of political conservatism and religiosity (H4). In line with our hypotheses, the results revealed a significant positive correlation between scores on the NJT-HH and the PSAM of honesty ($r = .40, p < .001$), and no significant correlations with scores on the other PSAMs.

Discussion study 2

The results of Study 2 provide some initial evidence that NJT-HH scores represent HH based on their relationships with scores on an explicit measure of the HEXACO traits (i.e., the HEXACO-60) and implicit measures of one similar construct (i.e., the PSAM of honesty) and two dissimilar constructs (i.e., the PSAMs of political conservatism and religiosity). In line with previous studies on implicit measures (e.g., Cunningham et al., 2001), NJT-HH scores correlated more strongly with scores on an implicit measure of a similar construct than with scores on an explicit measure of a similar construct. Importantly, the non-significant correlations between scores on the NJT-HH and the PSAMs of political conservatism and religiosity provide evidence that the NJT-HH scores assess personality trait differences rather than idiosyncratic response tendencies to vignettes (Hopkins & Wand, 2007; King & Wand, 2007). However, Study 1 did not allow testing the validity of the NJT-HH scores for predicting relevant external criteria. Hence, the goal of Study 3 is to provide further evidence for the NJT-HH scores' construct-related and criterion-related validity.

Study 3

The goal of Study 3 was to examine the NJT-HH scores' construct-related, criterion-related, and incremental validity by examining their relationship with HEXACO traits, CWB, and OCB. As we were only able to administer a limited number of NJT-HH items in this study, we see this study as a proof of concept that needs to be replicated with the full set of items.

Method study 3

Participants and procedure

The survey was distributed within the personal networks of three international master's students at a Dutch university (mostly family members, friends, and colleagues who all had paid and mostly white-collar jobs), via an online link and through paper-and-pencil administration. A power analysis (Faul et al., 2009) showed that the minimum required sample size was 155 to have 90% power to detect

a small- to medium-sized correlation between the NJT-HH and our external criteria ($r = .23$; based on Vargas et al., 2004), with $\alpha = .05$. The inclusion criteria for the study were being at least eighteen years old, having a paid job for at least one year with two working days a week, and having sufficient English language proficiency (measured through self-evaluations). As we expected that some participants might not meet these inclusion criteria, we oversampled participants to account for possible exclusions.

The survey consisted of four measures: the NJT-HH, the HEXACO-60 (De Vries, 2013; Lee & Ashton, 2004), a CWB measure (Spector et al., 2010), and an OCB measure (Kelloway et al., 2002; Smith et al., 1983). The study was approved by the faculty's ethics committee. Participants signed an informed consent that included information about their rights, the data protection procedures, and the researcher's contact information. At the end of the survey, participants were debriefed about the purpose of the study and were thanked for their contribution. Participants received a small gadget in return for their participation.

A total of 246 participants completed the survey. Sixteen participants were excluded from the analyses because they did not meet the job contract inclusion criterion or because they had finished the survey in an unrealistically short time (i.e., less than 500 seconds; median survey response time = 19.15 minutes). The sample ($M_{age} = 34.99$ years, $SD = 13.30$; 57.8% identified as men) consisted of 152 employees and 78 students who had a part-time job. Most participants had obtained a bachelor's degree ($n = 104$; 45.2%), a master's degree ($n = 48$; 20.9%), or a high school degree ($n = 35$; 15.2%). Most of the participants were born in the Netherlands ($n = 169$, 73.5%), and the other participants were born in other European countries ($n = 27$; 11.7%), Asia ($n = 27$; 11.7%), or another continent ($n = 10$; 4.3%). The participants had 1 to 45 years of work experience ($M = 14.81$, $SD = 11.75$), and worked 8 to 60 hours per week ($M = 32.75$, $SD = 10.93$). The participants who completed the online survey ($n = 152$) and the paper-and-pencil survey ($n = 78$) differed significantly in age (respectively $M = 36.72$, $SD = 14.53$ and $M = 31.63$, $SD = 9.74$; $t[212.17] = 3.15$, $p = .002$, $d = 0.41$) and working hours (respectively $M = 30.28$, $SD = 11.71$ and $M = 37.50$, $SD = 7.20$; $t[219.66] = -5.75$, $p < .001$, $d = 0.74$). However, controlling for these group differences in age and working hours did not affect the significance of any of the effects.

Materials

Normative Judgment Test of Honesty-Humility (NJT-HH)

The survey was part of a student project with a limited survey length (max. 20 minutes) to boost the overall response rate and minimize the negative consequences of nonresponse bias (Rogelberg & Stanton, 2007). For this reason, we selected 12 of the 23 developed items (9 of these items are also part of the 18-item NJT-HH). Specifically, to capture all HH aspects, we selected three items per HH facet that showed minimal overlap in content. The scores on these 12 items correlated at $r = .89$ ($p < .001$) with the scores on the 18-item NJT-HH in Study 2 and at $r = .88$ ($p < .001$) with the scores on the 18-item NJT-HH in Study 4. We conducted an EFA (with SPSS version 27) with the 12 items using the principal axis factoring extraction method. The KMO of .74 verified the sampling adequacy for the analysis and Bartlett's test of sphericity ($\chi^2 [66] = 330.62$, $p < .001$) indicated that the correlations between items were sufficiently large for the principal axis factoring method. The scree plot indicated one factor, which explained 23.49% variance, with factor loadings between .22 and .52. The alpha coefficient of the NJT-HH scores in the current study was .69.

HEXACO

We used 32 items of the 100-item HEXACO-PI-R (Lee & Ashton, 2004) to assess HH ($\alpha = .81$) and C ($\alpha = .82$), because these are integrity-related traits (Marcus et al., 2007) that are usually positively related to desirable employee behaviors and negatively related to undesirable employee behaviors (e.g., De Vries et al., 2017; Lee et al., 2005, 2019). The other four personality dimensions (E, X, A, and O) were measured with the Brief HEXACO Inventory (BHI; De Vries, 2013). This inventory includes four

items per dimension; each item belongs to a unique facet of the dimension. Example items are provided in Study 2. The alpha coefficients in the current sample were .49 for E, .60 for X, .34 for A, and .38 for O. These (very) low alpha coefficients are similar to previous studies (De Vries, 2013; Julian et al., 2022). It is important to note that there are many problems with using alpha as a reliability estimate for scores on short scales (e.g., Sijsma, 2009; Ziegler et al., 2014) as well as its contribution to score validity compared to other reliability estimates such as test-retest reliability (McCrae et al., 2011). As each scale item in the BHI represents a different facet, the items are broad in content and therefore less internally consistent (Julian et al., 2022). Indeed, despite these low alpha coefficients, research demonstrated that BHI scores have high temporal consistency (average 2-month test-retest stability is .76; De Vries, 2013), high convergence with scores on longer HEXACO measures (De Vries, 2013; Julian et al., 2022), and adequate criterion-related validity (e.g., Udayar et al., 2018).⁵ In both the 100-item HEXACO-PI-R and the BHI, items were rated on a five-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*).

Counterproductive Work Behavior (CWB)

We used the short 10-item version of the Counterproductive Work Behavior Checklist (CWB-C; Spector et al., 2010) to measure CWB. Example items are “Made fun of someone’s personal life” and “Started an argument with someone at work.” Participants indicated on a 5-point Likert scale how frequently they engaged in each behavior, ranging from 1 = *never* to 5 = *every day*. Coefficient alpha was .78.

Organizational Citizenship Behavior (OCB)

We measured OCB with nine items from Smith et al. (1983), see Kelloway et al. (2002). Example items are “Helping other employees with their work when they have been absent” and “Volunteering to do things not formally required by the job.” Participants indicated the extent to which each statement characterizes them, using a 5-point Likert scale that ranges from 1 = *not at all* to 5 = *extremely*. Coefficient alpha was .82.

Results study 3

We first compared the correlations of the scores on the NJT-HH and HEXACO HH with age, gender, and nationality. The correlation between HEXACO HH scores and age ($r = .42$, $p < .001$) was significantly stronger ($z = 3.73$, $p < .001$) than the correlation between the NJT-HH scores and age ($r = .14$, $p = .032$). Women ($M = 3.62$, $SD = 0.52$) had higher scores than men ($M = 3.32$, $SD = 0.57$) on HEXACO HH ($t[228] = -4.06$, $p < .001$, $d = 0.54$), while women ($M = 3.09$, $SD = 0.55$) and men ($M = 2.95$, $SD = 0.53$) had similar scores on the NJT-HH ($t[228] = -1.97$, $p = .050$, $d = 0.26$). Dutch participants ($M = 3.60$, $SD = 0.47$) had higher scores on HEXACO HH than non-Dutch participants ($M = 3.02$, $SD = 0.60$; unequal variance $t[87.68] = 6.83$, $p < .001$, Hedge’s $g = 1.14$ ⁶). However, Dutch participants ($M = 2.96$, $SD = 0.54$) had lower scores on the NJT-HH than non-Dutch participants ($M = 3.14$, $SD = 0.53$; $t[228] = -2.32$, $p = .021$, $d = 0.35$). The correlation between scores on the NJT-HH and CWB was not significantly different ($z = -0.07$, $p = .471$) between Dutch ($r = -.34$, $p < .001$) and non-Dutch ($r = -.33$, $p = .005$) participants. Similarly, the correlation between scores on the NJT-HH and OCB was not significantly different ($z = -0.33$, $p = .370$) between Dutch ($r = .21$, $p = .007$) and non-Dutch ($r = .26$, $p = .047$) participants.

Hypothesis testing

Means, standard deviations, and bivariate intercorrelations of the scores on the sociodemographic variables, the NJT-HH, the HEXACO traits, CWB, and OCB are presented in Table 4. We

⁵Accordingly, the correlations between scores on the BHI scales and the NJT-HH in Study 3 are quite similar to the correlations between scores on the longer HEXACO measures and the NJT-HH in Studies 2 and 4.

⁶We used Hedge’s g to account for the unequal variances in the Dutch and non-Dutch sample (see Delacre et al., 2017).

Table 4. Means, standard deviations, and bivariate intercorrelations of variables in study 3.

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. Gender	0.42	0.50	(-)											
2. Age	34.99	13.30	.08	(-)										
3. WH	32.75	10.93	-.29**	.26**	(-)									
4. NJT-HH	3.00	0.54	.13	.14*	.09	(.69)								
5. HH	3.45	0.57	.26**	.42*	-.01	.26**	(.81)							
6. E	2.89	0.68	.43**	-.03	-.20**	.12	-.03	(.49)						
7. X	3.80	0.65	.21**	.12	-.02	.03	.25**	-.09	(.60)					
8. A	2.97	0.57	-.04	-.05	-.04	.10	.07	-.08	-.07	(.34)				
9. C	3.56	0.51	.21**	-.02	.02	.09	.34**	-.03	.37**	-.02	(.82)			
10. O	3.55	0.58	-.03	.08	.04	-.07	.12	-.05	.24**	-.08	.18**	(.38)		
11. OCB	3.39	0.67	.03	-.09	.06	.21**	.09	-.12	.35**	.02	.40**	.08	(.82)	
12. CWB	1.70	0.49	-.16*	-.22**	.01	-.32**	-.41**	.06	-.13	-.13	-.31**	-.06	-.06	(.78)

Note. Gender: male = 0, female = 1, WH = Working hours, NJT-HH = Normative Judgment Test of Honesty-Humility, HH = Honesty-Humility, E = Emotionality, X = Extraversion, A = Agreeableness, C = Conscientiousness, O = Openness to Experience, OCB = Organizational citizenship behavior, CWB = Counterproductive work behavior. The possible range of scores for the scales is 1.00–5.00. *N* = 230.

p* < .05; *p* < .01 (two-tailed).

Table 5. Hierarchical regression analyses with HEXACO HH and the NJT-HH as predictors of CWB and OCB in study 3.

	CWB				OCB			
	Model 1		Model 2		Model 1		Model 2	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI
HH	-.41***	[-.53; -.29]	-.35***	[-.47; -.23]	.09	[-.04; .22]	.04	[-.10; .17]
NJT-HH			-.23***	[-.35; -.11]			.20**	[.06; .33]
<i>R</i> ²	.170		.219		.007		.043	
<i>F</i>	46.81***		31.90***		1.69		5.10**	
ΔR^2			.049				.036	
ΔF			14.26***				8.45**	

Note. HH = Honesty-Humility, OCB = Organizational citizenship behavior, NJT-HH = Normative Judgment Test of Honesty-Humility, CWB = Counterproductive work behavior. *N* = 230.

p* < .05; *p* < .01; ****p* < .001.

hypothesized that NJT-HH scores correlate modestly and positively with HEXACO HH scores (H1), but not with the scores on the other five HEXACO traits (H2). The results revealed a significant modest and positive correlation between NJT-HH and HEXACO HH scores ($r = .26$, $p < .001$), supporting H1. No significant correlations were found between NJT-HH scores and any other HEXACO trait scores, supporting H2. Furthermore, we predicted that NJT-HH scores are negatively correlated with CWB (H5) and positively correlated with OCB (H6). Indeed, the results revealed a negative correlation between NJT-HH scores and CWB ($r = -.32$, $p < .001$), and a positive correlation between NJT-HH scores and OCB ($r = .21$, $p = .002$), supporting H5 and H6.

We also formulated two hypotheses regarding the incremental validity of NJT-HH scores. We predicted that NJT-HH scores explain unique variance in CWB and OCB, above and beyond the variance explained by HEXACO HH scores (H7 and H8, respectively). To test these hypotheses, hierarchical regression analyses were conducted with CWB and OCB as the dependent variables (Table 5). HEXACO HH scores were included in the first step (Model 1), and NJT-HH scores were added in the second step (Model 2). In predicting CWB, HEXACO HH scores showed a significant negative beta weight in the first step ($\beta = -.41$, $t = -6.84$, $p < .001$), and NJT-HH scores showed a significant negative beta weight in the second step ($\beta = -.23$, $t = -3.78$, $p < .001$), supporting H7. Model 2 explained 21.9% of the variance in CWB ($F[2, 227] = 31.90$, $p < .001$), with 4.9% incremental variance explained compared to Model 1. Furthermore, in predicting OCB, there was a nonsignificant beta weight of HEXACO HH scores in the first step ($\beta = .09$, $t = 1.30$, $p = .195$). However, NJT-HH

scores showed a significant positive beta weight in the second step ($\beta = .20$, $t = 2.91$, $p = .004$), supporting H8. Model 2 explained 4.3% of the variance in OCB ($F[2, 225] = 5.10$, $p = .007$), that is, 3.6% incremental variance explained compared to Model 1.

We conducted additional analyses to test whether NJT-HH scores explain unique variance in CWB and OCB, above and beyond the variance explained by all HEXACO scores. The table with the results is included in the Supplementary Material (Table S1). In the hierarchical regression analyses, the six HEXACO scores were included in the first step (Model 1), and NJT-HH scores were added in the second step (Model 2). In predicting CWB, HEXACO HH scores ($\beta = -.34$, $t = -5.33$, $p < .001$) and HEXACO C scores ($\beta = -.21$, $t = -3.17$, $p = .002$) showed significant negative beta weights in the first step. This model explained 21.8% of the variance in CWB ($F[6, 223] = 10.34$, $p < .001$). Model 2 showed that NJT-HH scores were significantly negatively related to CWB ($\beta = -.23$, $t = -3.82$, $p < .001$), and explained 4.8% unique variance in CWB, above and beyond the variance explained by the six HEXACO scores ($\Delta F[1, 222] = 14.55$, $p < .001$). Furthermore, in a hierarchical regression analysis with OCB as the dependent variable, HEXACO X scores ($\beta = .26$, $t = 3.99$, $p < .001$) and HEXACO C scores ($\beta = .33$, $t = 5.06$, $p < .001$) were significantly positively related to OCB. This model explained 22.3% of the variance in OCB ($F[6, 221] = 10.73$, $p < .001$). Model 2 showed that NJT-HH scores were significantly positively related to OCB ($\beta = .22$, $t = 3.59$, $p < .001$), and explained 4.2% unique variance in OCB, above and beyond the variance explained by the six HEXACO scores ($\Delta F[1, 220] = 12.75$, $p < .001$).

Discussion study 3

The goal of Study 3 was to provide a proof of concept for the NJT-HH scores' construct-related, criterion-related, and incremental validity. In line with the hypotheses, the results showed that NJT-HH scores (a) are modestly and positively associated with HEXACO HH scores and not significantly associated with scores on the other five HEXACO traits, (b) are negatively associated with CWB and positively associated with OCB, and (c) explain unique variance in CWB and OCB, above and beyond the variance explained by HEXACO HH scores. Furthermore, the additional analyses showed that NJT-HH scores explained unique variance in CWB and OCB, above and beyond the variance explained by the six HEXACO trait scores. We also found that, with respect to age, gender, and nationality, score differences were significantly smaller on the NJT-HH than on the HEXACO HH scale. Altogether, these findings provide initial support for the criterion-related and incremental validity of NJT-HH scores.

However, the findings of Study 3 are subject to at least three limitations. First, the employee behaviors were measured using self-reports, which may have inflated the relationship between the predictors and the criteria due to common source bias (Meier & O'Toole, 2013; Podsakoff et al., 2003). Yet, as research shows that other-reports add little to the assessment of CWB to self-reports (Berry et al., 2012), this limitation pertains primarily to the measurement of OCB. Nonetheless, it is important to also examine the criterion-related validity of NJT-HH scores using other-reports of work outcomes (e.g., Jaramillo et al., 2005). Second, only a selection of NJT-HH items was used in this study, which may have reduced the reliability and, hence, the validity of the test scores. Third, four of the HEXACO traits were measured with the BHI. Although the validity of BHI scores has been demonstrated (De Vries, 2013), it is also important to test the construct-related and incremental validity of NJT-HH scores with a longer version of the HEXACO inventory.

In Study 4, we focused on CWB and OCB. Another important employee behavior is task performance, which has been defined as "activities that contribute to the organization's technical core either directly by implementing a part of its technological process, or indirectly by providing it with needed materials or services" (Borman & Motowidlo, 1997, p. 99). Whereas a conceptual and empirical link could be made between HH and CWB and OCB, the link between HH and task performance is less evident (Lee et al., 2019): Employees high in HH are sincere, fair, modest, and avoid greed (Ashton &

Lee, 2007). These characteristics do not always provide a clear advantage in carrying out activities that contribute to the organization's technical core (with caregiving organizations as a notable exception; Johnson et al., 2011). However, due to globalization, demographic shifts, and rapid technological advancements, more and more jobs are placing an emphasis on HH-related constructs such as cooperation and integrity (Anglim et al., 2019; Robles, 2012). Hence, HH may be a more important predictor of task performance than the current literature suggests.

So far, only a few studies have examined HH as a predictor of task performance. In a recent meta-analysis with seven studies, a weak positive relationship was found between HH and task performance (i.e., $r = .10$; Lee et al., 2019). However, some studies revealed no effect (e.g., Oh et al., 2014). Furthermore, there is no consistent evidence that dark triad traits – traits that are associated with callous, selfish, and malevolent interpersonal behaviors (Paulhus & Williams, 2002), and that have high conceptual similarities with the opposite end of HH (e.g., Hodson et al., 2018) – are related to task performance (O'Boyle et al., 2012). Altogether, there is currently only indirect and inconsistent support for a relationship between HH and task performance. Therefore, we propose two research questions for the relationship between NJT-HH scores and task performance:

Research question 1 (RQ1): To what extent are NJT-HH scores correlated with task performance?

Research question 2 (RQ2): To what extent do NJT-HH scores explain unique variance task performance, above and beyond the variance explained by HEXACO HH scores?

In addition to exploratorily examining the relationship between NJT-HH scores and task performance, another goal of Study 4 was to provide an additional test of our hypotheses based on self-ratings of CWB and supervisor ratings of CWB and OCB. In the analyses with supervisor ratings, we controlled for supervisor-subordinate interaction frequency, because this could influence supervisor ratings (Kacmar et al., 2003). Furthermore, to deal with some of the limitations of Study 3, we administered the full NJT-HH and the HEXACO-60 inventory (Ashton & Lee, 2009), and we used longer and more common scales of CWB and OCB.

Method study 4

Participants

The participants in the present study were employees and their supervisors, typically in white-collar jobs. A power analysis (Faul et al., 2009) showed that the minimum required sample size was 111 to have 90% power to replicate the average criterion-related validity coefficient in Study 3 ($r = .27$), with $\alpha = .05$. Employees had to meet four criteria to participate in this study: being between the age of 18 and 65 years, being born in the Netherlands, having worked for the current company for at least six months with a minimum of 16 hours per week, and having sufficient English proficiency to understand the survey. Supervisors also had to meet four criteria to participate in this study: being between the age of 18 and 65 years, having supervised the employee for a minimum of three months, having interacted with the employee at least once a week, and having sufficient English proficiency to understand the survey (measured through self-evaluations).

The final dataset consisted of 123 employees (63.4% identified as men) and 93 employee-supervisor dyads (59 employees and 55 supervisors identified as men). On average, the employees were 32.89 years old ($SD = 10.49$). Most employees had obtained a bachelor's degree ($n = 72$; 58.5%), a master's degree ($n = 22$; 17.9%), or an associate degree ($n = 21$; 17.1%). Most of the employees (100; 81.3%) had a permanent contract. The employees had 2 to 45 years of work experience ($M = 13.90$, $SD = 10.81$), and had an average organizational tenure of 3.80 years ($SD = 5.28$), with working hours ranging from 16 to 60 hours per week ($M = 38.11$, $SD = 8.16$). The supervisors were on average 38.52 years old ($SD = 8.62$). Most supervisors had obtained a bachelor's degree ($n = 45$; 48.4%) or a master's degree ($n = 37$; 39.8%). The average supervisors' organizational tenure was 6.84 years ($SD = 6.99$), and their average

managerial experience was 7.55 years ($SD = 6.92$). The supervisors worked on average 43.26 hours per week ($SD = 9.44$), varying from 24 to 70 hours. The employee-supervisor dyads worked together on average for 1.88 years ($SD = 2.35$).

Procedure

The data were collected in the Netherlands. The employee-supervisor dyads were recruited by three master's students who worked collaboratively on a research project. Participants were recruited through the students' internship providers and through a few other organizations located in the Netherlands. First, the students contacted potential supervisors, explained the aim of the research, and invited them to participate. If supervisors were willing to participate in this study, they were asked to provide the names of their subordinates who met the participation criteria. Second, the students contacted these employees, explained the aim of the research, and asked them to participate in this study. The employees were asked to complete the survey that included the NJT-HH, the HEXACO, and a scale for CWB. At the end of the survey, the employees had to enter a random 6-digit code. The employees were asked to share this code with their supervisor, together with the supervisor's survey link that was provided at the end of their survey. Subsequently, at the beginning of the supervisor's survey, the supervisors entered the 6-digit code. This survey included a scale for CWB, OCB, and task performance. This procedure enabled us to match the surveys of the employees and the supervisors while ensuring participants' anonymity and the confidentiality of their responses. The present research was approved by the faculty's ethics committee. In the consent form, participants were informed about their rights and the anonymous data processing and the confidential treatment of their data for scientific research. In the debriefing of the surveys, the goal of the study was explained, and the participants were thanked for their participation. The employee survey took about 15–20 minutes to complete, and the supervisor survey took about 10–15 minutes to complete.

Materials

Normative Judgment Test of Honesty-Humility (NJT-HH)

We administered the 18-item NJT-HH. Again, we conducted an EFA (with SPSS version 27) using the principal axis factoring extraction method ($KMO = .75$; Bartlett's test of sphericity = 470.23, $p < .001$). The scree plot showed an inflexion that would justify retaining one factor, which explained 23.13% variance. The factor loadings varied between .24 and .61. The alpha coefficient in the current study was .79.

HEXACO

We used the HEXACO-60 (Ashton & Lee, 2009; De Vries et al., 2009) to measure the six HEXACO traits among employees. Items were rated on a 5-point Likert scale (1 = *strongly disagree*; 5 = *strongly agree*). Example items are provided in Study 2. The alpha coefficients in the current study were .71 for HH, .69 for E, .80 for X, .67 for A, .79 for C, and .73 for O.

Counterproductive Work Behavior (CWB)

The 19-item CWB scale by Bennett and Robinson (2000) was administered to both the employees and their supervisors. Example items are "Made fun of someone at work" and "Taken property without permission." Items were rated on a 5-point Likert scale, ranging from 1 = *never* to 5 = *every day*. Coefficient alpha was .83 for both the self-reports and the supervisory ratings.

Organizational Citizenship Behavior (OCB)

We measured supervisor ratings of OCB with the 16-item OCB scale by Lee and Allen (2002). Supervisors were asked to indicate how often their subordinate(s) engaged in certain behaviors,

ranging from 1 = *never* to 5 = *always*. Example items are “Defend the organization when other employees criticize it” and “Willingly give their time to help others who have work-related problems.” Coefficient alpha was .89.

Task Performance

To measure supervisor ratings of task performance, we used nine items from Goodman and Svyantek (1999). Example items are “Achieves the objective of the job” and “Demonstrates expertise in all job-related tasks.” Items were rated on a 5-point Likert scale, ranging from 1 = *strongly disagree* to 5 = *strongly agree*. Coefficient alpha was .86.

Interaction Frequency

Supervisor ratings are affected by the frequency of interaction between the supervisor and their subordinate (Kacmar et al., 2003). To take this potential confound into consideration, supervisors completed a 4-item scale by McAllister (1995) that measures the employee-supervisor interaction frequency. An example item of this scale is “How frequently do you interact with this person at work informally or socially?.” Response alternatives ranged from 1 = *once or twice in the last 6 months* to 7 = *many times daily*. Higher scores on this scale indicate perceptions of a higher communication frequency. Coefficient alpha was .87.

Results study 4

In the current study, there were no significant differences between men and women in NJT-HH scores ($M = 3.65$, $SD = 0.47$ and $M = 3.62$, $SD = 0.46$, respectively; $t[119] = 0.28$, $p = .781$, $d = 0.05$), HEXACO HH scores ($M = 3.51$, $SD = 0.58$ and $M = 3.59$, $SD = 0.56$, respectively; $t[118] = -0.67$, $p = .506$, $d = 0.13$), or any other HEXACO scores. The only significant difference between men ($M = 40.39$, $SD = 7.07$) and women ($M = 34.00$, $SD = 8.64$) was their working hours per week, unequal variance $t(73.26) = 4.14$, $p < .001$, Hedge’s $g = 0.83$. Furthermore, employees’ age was positively correlated with NJT-HH scores ($r = .28$, $p = .002$) and HEXACO HH scores ($r = .36$, $p < .001$), and these correlations did not differ significantly from each other ($z = 0.80$, $p = .212$). Age was negatively correlated with supervisor ratings of CWB ($r = -.24$, $p = .024$) and positively correlated with their ratings of OCB ($r = .30$, $p = .004$). Age was not significantly correlated with self-ratings of CWB ($r = -.17$, $p = .067$), nor with supervisor ratings of task performance ($r = -.06$, $p = .593$). Finally, interaction frequency was negatively correlated with HEXACO C scores ($r = -.21$, $p = .041$), and positively correlated with supervisor ratings of CWB ($r = .25$, $p = .015$).

Hypothesis testing

Means, standard deviations, and bivariate intercorrelations of scores on the sociodemographic variables, the NJT-HH, the HEXACO traits, and the employee behaviors are presented in Table 6. In line with Hypothesis 1, the results revealed a significant modest positive correlation between the NJT-HH and HEXACO HH scores ($r = .41$, $p < .001$). Hypothesis 2 was largely supported, as we only found a modest and positive correlation between the NJT-HH and HEXACO C scores ($r = .18$, $p = .042$), and no significant correlations between the NJT-HH scores and any of the other HEXACO scores. Furthermore, we predicted that NJT-HH scores negatively correlated with CWB (H5). We found a correlation of $r = -.30$ ($p < .001$) between NJT-HH scores and self-ratings of CWB, and a correlation of $r = -.41$ ($p < .001$) between NJT-HH scores and supervisor ratings of CWB, supporting H5. We also predicted that NJT-HH scores are positively correlated with OCB (H6). Indeed, there was a positive correlation between NJT-HH scores and supervisor ratings of OCB ($r = .48$, $p < .001$), supporting H6.

We expected that NJT-HH scores explain unique variance in CWB (H7) and OCB (H8), above and beyond the variance explained by HEXACO HH scores. To test these hypotheses, hierarchical

Table 6. Means, standard deviations, and bivariate intercorrelations of variables in study 4.

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Gender	0.36	0.48	-														
2. Age	32.89	10.49	-.02	-													
3. WH	38.11	8.16	-.37**	-.11	-												
4. IF	5.21	1.10	-.05	-.16	.16	(.88)											
5. NJT-HH	3.64	0.47	-.03	.34**	.15	-.02	(.79)										
6. HH	3.55	0.57	.06	.36**	.08	-.11	.41**	(.71)									
7. E	2.96	0.52	.50**	-.02	-.10	.03	.02	.01	(.69)								
8. X	3.61	0.56	-.09	-.05	.19*	.19	.15	.07	-.11	(.80)							
9. A	3.13	0.49	-.04	.06	.04	.01	.13	.33**	-.02	.01	(.67)						
10. C	3.63	0.54	.13	.03	-.00	-.21*	.18*	.35**	.09	.09	.05	(.79)					
11. O	3.41	0.57	-.07	-.08	.11	-.18	-.03	-.02	-.09	.05	.19*	-.03	(.73)				
12. OCB	3.53	0.57	-.06	.30**	.25*	.11	.48**	.37**	.09	.24*	.15	.17*	-.01	(.89)			
13. CWB-E	1.51	0.36	-.07	-.17	.08	.13	-.30**	-.41**	.00	-.15	-.20*	-.28**	.12	-.12	(.83)		
14. CWB-S	1.28	0.29	.04	-.24*	.03	.25*	-.41**	-.45**	.05	-.21*	-.11	-.30**	-.24*	-.42**	.46**	(.83)	
15. TP	3.87	0.50	.11	-.06	.07	-.03	.23*	.04	.10	.10	.05	.18	.09	.43**	.03	-.33**	(.86)

Note. Gender: male = 0, female = 1, WH = Working hours, IF = Interaction Frequency, NJT-HH = Normative Judgment Test of Honesty-Humility, HH = Honesty-Humility, E = Emotionality, X = Extraversion, A = Agreeableness, C = Conscientiousness, O = Openness to Experience, OCB = Organizational citizenship behavior, CWB-E = Counterproductive work behavior employee self-reports, CWB-S = Counterproductive work behavior supervisory rating, TP = Task performance. The possible range of scores for the scales is 1.00–5.00, except for IF, which has a possible range of 1.00–7.00. N=93 to N=123.

* $p < .05$; ** $p < .01$ (two-tailed).

Table 7. Hierarchical regression analyses with predictors of CWB and OCB in study 4.

	Model 1		Model 2	
	β	95% CI	β	95% CI
CWB-E		$R^2 = .17$		$R^2 = .19$ ($\Delta R^2 = .02$)
H	-.41***	[-.57; -.24]	-.34***	[-.52; -.17]
NJT-HH			-.16	[-.34; .02]

	Model 1		Model 2		Model 3	
	β	95% CI	β	95% CI	β	95% CI
CWB-S		$R^2 = .06$		$R^2 = .24$ ($\Delta R^2 = .18$)		$R^2 = .30$ ($\Delta R^2 = .06$)
IF	.25*	[.05; .45]	.21*	[.02; .39]	.21*	[.03; .39]
H			-.43***	[-.61; -.24]	-.31**	[-.50; -.12]
NJT-HH					-.28**	[-.47; -.08]
OCB		$R^2 = .01$		$R^2 = .16$ ($\Delta R^2 = .15$)		$R^2 = .28$ ($\Delta R^2 = .12$)
IF	.11	[-.10; .32]	.15	[-.04; .35]	.14	[-.04; .32]
H			.39***	[.19; .58]	.23*	[.03; .42]
NJT-HH					.39***	[.19; .58]
TP		$R^2 = .00$		$R^2 = .00$ ($\Delta R^2 = .00$)		$R^2 = .06$ ($\Delta R^2 = .05$)
IF	-.03	[-.23; .18]	-.02	[-.23; .19]	-.03	[-.23; .18]
H			.04	[-.17; .25]	-.07	[-.29; .15]
NJT-HH					.26*	[.03; .48]

Note. IF = Interaction Frequency, HH = Honesty-Humility, NJT-HH = Normative Judgment Test of Honesty-Humility, CWB-E = Counterproductive work behavior employee self-reports, CWB-S = Counterproductive work behavior supervisory rating, OCB = Organizational citizenship behavior. $N = 123$ and $N = 93$.

* $p < .05$; ** $p < .01$; *** $p < .001$.

regression analyses were conducted with self-reported CWB and supervisor ratings of CWB and OCB as the dependent variables (Table 7). For self-ratings of CWB, HEXACO HH scores were included in the first step (Model 1), and NJT-HH scores were added in the second step (Model 2). For supervisor ratings of CWB and OCB, interaction frequency was included as a control variable in the first step (Model 1), HEXACO HH scores were included in the second step (Model 2), and NJT-HH scores were added in the third step (Model 3). In the Supplementary Materials, we present the results of the hierarchical regression analysis without the control variable interaction frequency (Table S2).

In the hierarchical regression analysis with self-ratings of CWB as the dependent variable, HEXACO HH scores showed a negative beta weight in Model 1 ($\beta = -.41$, $t = -4.91$, $p < .001$), explaining 16.8% of the variance in CWB self-ratings ($F[1, 120] = 24.15$, $p < .001$). Model 2 showed that NJT-HH scores ($\beta = -.16$, $t = -1.75$, $p = .083$) did not explain unique variance in self-ratings of CWB above and beyond the variance explained by HEXACO HH scores ($F[2, 119] = 13.81$, $p < .001$). In the hierarchical regression analysis with supervisor ratings of CWB as the dependent variable, interaction frequency ($\beta = .25$, $t = 2.47$, $p = .015$) explained 6.3% variance ($F[1, 91] = 6.11$, $p = .015$). In Model 2, HEXACO HH scores ($\beta = -.43$, $t = -4.60$, $p < .001$) showed a significant negative beta weight. This model explained 24.1% of the variance in supervisor ratings of CWB ($F[2, 90] = 14.30$, $p < .001$). Model 3 showed that NJT-HH scores were significantly negatively related to supervisor ratings of CWB ($\beta = -.28$, $t = -2.85$, $p = .005$), and explained 30.5% variance in supervisor ratings of CWB ($F[3, 89] = 13.00$, $p < .001$), that is, 6.3% unique variance above and beyond the variance explained by Model 2. Altogether, these results provide partial support for H7.

In the hierarchical regression analysis with supervisor ratings of OCB as the dependent variable, interaction frequency ($\beta = .11$, $t = 1.06$, $p = .290$) showed no significant beta weight. In Model 2, HEXACO HH scores ($\beta = .39$, $t = 3.99$, $p < .001$) showed a significant and positive beta weight. This model explained 16.0% of the variance in supervisor ratings of OCB ($F[2, 90] = 8.60$, $p < .001$). Model 3 showed that NJT-HH scores were significantly and positively related to supervisor ratings of OCB ($\beta = .39$, $t = 3.93$, $p < .001$). This model explained 28.4% variance in supervisor ratings of OCB ($F[3, 89] = 11.79$, $p < .001$), which is 12.4% additional variance above and beyond the variance explained by Model 2. Altogether, these results provide support for H8.

Table 8. An overview of the hypotheses, research questions, and their empirical support.

#	Hypothesis or research question	Support
H1	The NJT-HH is modestly and positively correlated with HEXACO HH.	Yes
H2	The NJT-HH is not significantly correlated with any of the five other HEXACO traits.	Largely supported (weak correlation with Agreeableness in Study 2 and Conscientiousness in Study 4)
H3	The NJT-HH is modestly and positively correlated with the PSAM of honesty.	Yes
H4	The NJT-HH is not significantly correlated with the PSAM of political conservatism and religiosity.	Yes
H5	The NJT-HH is negatively correlated with CWB.	Yes
H6	The NJT-HH is negatively correlated with OCB.	Yes
H7	The NJT-HH explains unique variance in CWB above the variance explained by HEXACO HH	Yes
H8	The NJT-HH explains unique variance in OCB above the variance explained by HEXACO HH.	Yes
RQ1	To what extent is the NJT-HH correlated with task performance?	$r = .23$
RQ2	To what extent does the NJT-HH explain unique variance task performance, above and beyond the variance explained by HEXACO HH?	5.4% unique variance explained

We conducted additional analyses to test whether NJT-HH scores explain unique variance in CWB and OCB, above and beyond the variance explained by all six HEXACO trait scores. The table of the results is included in the Supplementary Material (Table S3). In the hierarchical regression analyses, interaction frequency was included in the first step (Model 1), the six HEXACO scores were included in the second step (Model 2), and NJT-HH scores added in the third step (Model 3). Step 1 was not applied in the analyses with CWB self-reports. The results showed that in predicting CWB self-reports, NJT-HH scores did not explain unique variance in CWB self-reports above and beyond the variance explained by the six HEXACO scores ($\beta = -.13$, $t = -1.47$, $p = .144$). However, in predicting supervisor ratings of CWB, NJT-HH scores ($\beta = -.25$, $t = -2.69$, $p = .009$) explained 5.2% unique variance above and beyond the variance explained by the six HEXACO scores ($\Delta F[1, 84] = 7.24$, $p = .009$). Finally, in predicting supervisor ratings of OCB, NJT-HH scores ($\beta = .36$, $t = 3.60$, $p < .001$) explained 10.6% unique variance in supervisor ratings of OCB ($\Delta F[1, 84] = 12.97$, $p < .001$). These additional analyses showed that NJT-HH scores also explained unique variance in work outcomes above and beyond all HEXACO scores.

Exploratory research questions

We posited two exploratory research questions for the relations between NJT-HH scores and task performance. In RQ1, we proposed examining to what extent NJT-HH scores are correlated with task performance. The results showed that NJT-HH scores are significantly and positively correlated with task performance, $r = .23$ ($p = .028$). In RQ2, we proposed examining to what extent NJT-HH scores explain unique variance in task performance, above and beyond the variance explained by HEXACO HH scores (Table 7). After controlling for interaction frequency ($\beta = -.03$, $t = -0.25$, $p = .805$; Step 1), HEXACO HH scores showed no significant beta weight for supervisor ratings of task performance ($\beta = .04$, $t = 0.35$, $p = .724$). However, NJT-HH scores ($\beta = .26$, $t = 2.27$, $p = .026$) did show a significant and positive beta weight in predicting supervisor ratings of task performance, explaining 5.4% additional variance above and beyond HEXACO HH scores ($\Delta F[1, 89] = 5.13$, $p = .026$). Additionally, NJT-HH scores ($\beta = .24$, $t = 2.08$, $p = .040$) also explained 4.6% unique variance in task performance above and beyond the variance explained by the six HEXACO trait scores ($\Delta F[1, 84] = 4.33$, $p = .040$; see Supplementary Material, Table S4). Altogether, these exploratory research analyses showed that NJT-HH scores are positively related to task performance, and explain unique variance in task performance, above and beyond the variance explained by HEXACO HH scores as well as all six HEXACO trait scores.

Discussion study 4

Overall, Study 4 provides additional support for NJT-HH scores' construct- and criterion-related validity (for an overview of the results of Studies 2–4, see [Table 8](#)). In line with our expectations, NJT-HH scores are modestly and positively associated with HEXACO HH scores. Furthermore, NJT-HH scores show no significant correlations with the other HEXACO trait scores. NJT-HH scores are negatively associated with self-ratings and supervisor ratings of CWB and positively associated with supervisor ratings of OCB, and explain unique variance in the supervisor ratings of CWB and OCB, above and beyond the variance explained by HEXACO HH scores. Additional analyses showed that NJT-HH scores also explain unique variance in supervisor ratings of CWB and OCB, above and beyond the variance explained by the six HEXACO trait scores. Lastly, the exploratory research analyses showed that NJT-HH scores are positively associated with task performance, and explain unique variance in task performance above and beyond the variance explained by HEXACO HH scores and even all six HEXACO trait scores.

General discussion

The present research addresses recent calls to examine and advance alternative personality assessment methods, such as implicit instruments, for organizational contexts (Funder, 2002; Sackett et al., 2017). Here, we have developed the NJT-HH, an implicit instrument of Honesty-Humility (HH) that is based on the PSAM paradigm of Vargas et al. (2004). In four studies, we assessed the reliability of the NJT-HH scores, examined their factor structure, and investigated their relationships with scores on the HEXACO scales, PSAMs of honesty, political conservatism, and religiosity, as well as self-reported and supervisor ratings of employees' CWB, OCB, and task performance. The findings provide initial support for the NJT-HH scores' construct-related validity, criterion-related validity, and incremental validity in predicting employee behaviors.

Theoretical implications

The present study contributes to the literature on implicit instruments of personality (Back et al., 2009; Bosson et al., 2000; Hofmann et al., 2005; James et al., 2005) by validating an implicit instrument of HH based on the PSAM paradigm. While Vargas et al. (2004) provided a proof of concept for the PSAM, our research with the NJT-HH demonstrates that this paradigm is useful for assessing trait HH in organizational contexts. Specifically, the current research provides support for the construct-related validity of NJT-HH scores (Studies 2–4). We found a modest and positive relationship between scores on the NJT-HH and HEXACO HH, and no significant correlations between scores on the NJT-HH and the other HEXACO traits (apart from a weak, significant correlation with HEXACO A in Study 2 and HEXACO C in Study 4). In line with previous studies on implicit instruments (e.g., Back et al., 2009; James et al., 2005), we treated a modest positive correlation between scores on the implicit and explicit instrument of the same trait as evidence for convergent validity. In line with this interpretation, we argue that it is important to develop test validation guidelines for scores on implicit instruments that acknowledge modest correlations with scores on explicit instruments of the same construct as evidence for their convergent validity. Furthermore, we found a modest and positive correlation between scores on the NJT-HH and the PSAM of honesty, and no significant correlations between scores on the NJT-HH and the PSAMs of political conservatism and religiosity, which provides further evidence that the NJT-HH scores represent HH and not response tendencies to vignettes (Hopkins & King, 2007; King & Wand, 2007).

Moreover, the current research contributes to the literature on employee work behavior by revealing that NJT-HH scores are able to predict CWB, OCB, and task performance, and also able to explain unique variance in these employee behaviors above and beyond the variance explained by HEXACO HH scores, as well as all six HEXACO trait scores. Thus, the NJT-HH assesses unique

variance in one's personality, and adding the NJT-HH to a personality self-report measure makes it possible to more accurately predict employees' CWB, OCB, and task performance. This is an important and promising finding, as previous research has shown that only trait Conscientiousness is a consistent predictor of these three employee behaviors (Connelly & Ones, 2010; Lee et al., 2019). This suggests the potential for an NJT of Conscientiousness to likewise provide additional predictive power over and above self-rated Conscientiousness.

Furthermore, in line with previous studies showing no or only limited support for a relationship between HH and task performance (e.g., Lee et al., 2019; Oh et al., 2014), HEXACO HH scores were not significantly related to supervisory ratings of task performance. However, NJT-HH scores were positively related to task performance. We suspect that NJT-HH scores may be a more effective predictor of task performance because the NJT-HH items are more closely aligned to work contexts than those of the HEXACO, and because implicit measures are less prone to (un)intentional response distortion (e.g., by reporting ideal perceptions rather than actual ones; Fazio & Olson, 2003) compared to explicit measures. It could also be that the unique part of HH outside of employees' awareness provides a stronger advantage in carrying out activities that contribute to the organization's technical core than the part within employees' awareness. However, as the conceptual link between HH and task performance is less evident than the link between HH and CWB and OCB, we did not formulate specific predictions regarding these relationships. Hence, more conceptual and empirical work is needed to fully understand the nature of the HH-task performance link.

Finally, although it was not the primary goal of our research, the NJT-HH may address the problem that scores on more valid selection instruments show larger (particularly ethnic) subgroup differences (Ployhart & Holtz, 2008). Reducing these score differences is an important concern for organizations because of their potential influence on workforce diversity (e.g., Sackett et al., 2001). Importantly, integrity scores, which are highly related to HH scores (e.g., Lee et al., 2005, 2019; Marcus et al., 2007), are generally lower among individuals from collectivist cultures (Fine, 2010). Furthermore, in the HEXACO model, HH is one of the six traits that shows the largest gender differences after Emotionality (Lee & Ashton, 2020), and HH scores substantially increase with age (Ashton & Lee, 2016). In the present research, we showed that with respect to gender, age, and nationality, the score differences on the NJT-HH between demographic groups were small or nonexistent, and smaller than those of the HEXACO HH scale. Although these findings are promising, they need to be interpreted with caution as these score differences were examined in selective samples, recruited within the research team's personal networks. Hence, we encourage future studies to test the robustness of these findings in larger and more representative samples.

Practical implications

The NJT-HH is an instrument that is easy to administer and cost-efficient compared to other implicit instruments that require specialized test administration expertise, one-on-one administration, and complex scoring procedures (Bing et al., 2007; Lilienfeld et al., 2000). A "low tech" implicit instrument (Vargas et al., 2007) such as the NJT-HH could therefore be a feasible alternative or complement to self-report measures in organizational contexts. Personnel selection could be one useful application of the NJT-HH. The NJT-HH might also be a useful instrument for coaching and employee development. The use of personality assessments in employees' development has increased significantly in the last two decades (McDowall & Redman, 2017; Passmore, 2012). The goal of such development programs is to increase self-awareness (Cseh et al., 2013), which positively influences employees' well-being (e.g., Harrington & Loffredo, 2011). Importantly, people's self-knowledge is the lowest for

evaluative traits (e.g., Vazire & Carlson, 2010), and in particular trait HH (Thielmann et al., 2021). The present research provides some evidence that the HEXACO HH scale should be supplemented by the NJT-HH to obtain a more detailed picture of someone's level of HH. Together, these test scores could be used to set goals for self-development.

Limitations and future directions

The first limitation of this research is the investigation of NJT-HH scores in a low-stakes context. In a high-stakes context, job applicants may be inclined to substantially elevate their scores on self-report personality measures (e.g., Goffin & Christiansen, 2003; Griffith et al., 2007), and they do so mainly for socially desirable traits such as HH (Anglim et al., 2017). Some studies have empirically investigated the fakability of implicit instruments, and these studies usually show that these instruments are resistant to faking as long as the participant has not been informed about the construct being measured (LeBreton et al., 2007 for the CRT; Steffens, 2004 for the IAT). Thus, one important avenue for future research is to investigate the fakability of NJT-HH scores and compare them to the fakability of scores on self-report personality measures and other implicit instruments. Future research should also test the criterion-related validity of NJT-HH scores among a large sample of actual job applicants using predictive designs.

Notably, there was some inconsistency in the NJT-HH scores' model fit indices. Specifically, although the RMSEA indicated reasonable fit (e.g., MacCallum et al., 1996), the CFI and TLI values did not reach such thresholds (Study 1). The low CFI and TLI values are probably due to the relatively low covariances between item scores, which is quite common in tests with situational stems, including situational judgment tests (Kasten & Freund, 2015), which show similar CFA results (e.g., Oostrom et al., 2019). Although this finding warrants further examination using larger samples, the CFI and TLI values may not be a cause for alarm as the RMSEA is often considered "one of the most informative criteria in structural equation modeling" (Byrne, 1998, p. 112). Moreover, the NJT-HH scores showed promising psychometric properties in other domains (i.e., reliability, construct-related validity, and criterion-related validity).

Although the present study provided support for the NJT-HH scores' construct-related validity based on their relationships with scores on an explicit measure of the HEXACO traits and implicit measures of (dis)similar constructs based on the same paradigm, it is still worth examining the NJT-HH scores' construct-related validity in a broader nomological net of related constructs. A particularly relevant trait to study in this context is emotional intelligence. In the PSAM paradigm, participants need to judge others, and they do so by comparing the other to themselves (Dunning & Hayes, 1996). This comparison requires self-reflection, which is a crucial aspect of emotional intelligence (e.g., Boyatzis et al., 2000). Furthermore, some research has examined anchoring vignettes (comparable to the NJT-HH items) as a tool to assess response styles, aiming to improve the cross-cultural validity of instruments by controlling for the vignette scores (King & Wand, 2007; King et al., 2004). There is some evidence that controlling for anchoring vignette response style results in improvements in the validity of test scores, although the effects are usually weak (e.g., He et al., 2017; Primi et al., 2016). Although the convergent and discriminant correlations with the PSAM scores provide support that the instrument assesses a substantive construct and not (merely) response tendencies, we encourage future research to further examine the extent to which NJT-HH scores represent HH versus response tendencies. One methodology for this is to develop an NJT for a construct that is unrelated to HH (e.g., Extraversion; Thielmann et al., 2021), and investigate its relationship with NJT-HH scores.

Another opportunity for future research is to investigate applicant reactions (e.g., perceived fairness, liking; Ryan & Ployhart, 2000) to the NJT-HH. Applicant reactions to selection instruments are essential to study, as these reactions affect applicants' test performance, perceptions of organizational attractiveness, and intentions to accept a job offer (McCarthy et al., 2017). The available empirical work on applicant reactions to implicit instruments in the selection context has raised critical issues.

For example, research has shown that using an IAT, developed to predict training skills, to hire or promote individuals results in negative procedural justice reactions (Wright & Meade, 2011). For TAT-like instruments, concerns have been raised about their lack of face validity (Van Rensburg et al., 2019), which could lead to defensive test-taker responses (Ridgeway, 2017). To our knowledge, applicant reactions to CRTs have not been empirically investigated so far (Connelly et al., 2018). Future research needs to examine applicant reactions to the NJT-HH, and compare them to applicant reactions to personality self-reports and other implicit instruments.

Finally, future research could also investigate the cross-cultural validity of NJT-HH scores. One research suggestion is to test the criterion-related validity of NJT-HH scores in collectivistic cultures, where people have a more interdependent (versus independent) self-concept (Markus & Kitayama, 1991). Scholars have argued that an interdependent self-concept might contribute to lower validities of personality self-reports among people from collectivistic cultures, although evidence for this argument is inconsistent (Church & Katigbak, 2017). Arguably, NJT-HH scores might have higher criterion-related validity than self-reports of HH in collectivistic cultures, because self-reflection is crucial in personality self-reports but not for the NJT-HH. However, this is a prediction that needs to be tested in future research.

Conclusion

The present research provides initial evidence for the construct-related validity of NJT-HH scores, and shows that NJT-HH scores are able to predict CWB, OCB, and task performance. NJT-HH scores are also able to explain unique variance in these employee behaviors, above the variance explained by HEXACO HH scores as well as all six HEXACO trait scores. While research is necessary to provide more insights into the practical value of the NJT-HH (determined by, for instance, its fakability and applicant reactions), the present research suggests that implicit personality measures such as the NJT-HH could form a useful alternative or complement to personality self-reports in organizational contexts.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Stichting NOA .

Data availability statement

The data, syntax, and further information about the Materials of this study are available in <https://osf.io/t32m4/>

References

- Anglim, J., Morse, G., De Vries, R. E., MacCann, C., Marty, A., & Möttus, R. (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>
- Anglim, J., Sojo, V., Ashford, L. J., Newman, A., & Marty, A. (2019). Predicting employee attitudes to workplace diversity from personality, values, and cognitive ability. *Journal of Research in Personality*, 83, 103865. <https://doi.org/10.1016/j.jrp.2019.103865>
- Apers, C., Lang, J. W., & Deros, E. (2019). Who earns more? Explicit traits, implicit motives and income growth trajectories. *Journal of Vocational Behavior*, 110, 214–228. <https://doi.org/10.1016/j.jvb.2018.12.004>
- 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. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, 91(4), 340–345. <https://doi.org/10.1080/00223890902935878>
- Ashton, M. C., & Lee, K. (2016). Age trends in HEXACO-PI-R self-reports. *Journal of Research in Personality*, 64, 102–111. <https://doi.org/10.1016/j.jrp.2016.08.008>
- Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO model of personality structure—and why those objections fail. *European Journal of Personality*, 34(4), 492–510. <https://doi.org/10.1002/per.2242>
- Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO honesty–humility, agreeableness, and emotionality factors. *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., DiBlas, L., Boies, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, 86(2), 356–366. <https://doi.org/10.1037/0022-3514.86.2.356>
- Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology*, 97(3), 533–548. <https://doi.org/10.1037/a0016229>
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54(7), 462–479. <https://doi.org/10.1037/0003-066x.54.7.462>
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44(1), 1–26. <https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>
- Beauregard, K. S., & Dunning, D. (1998). Turning up the contrast: Self-enhancement motives prompt egocentric contrast effects in social judgments. *Journal of Personality and Social Psychology*, 74(3), 606–621. <https://doi.org/10.1037/0022-3514.74.3.606>
- 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>
- Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2012). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology*, 97(3), 613–636. <https://doi.org/10.1037/a0026739>
- Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, 92(2), 410–424. <https://doi.org/10.1037/0021-9010.92.2.410>
- Bing, M. N., Stewart, S. M., Davison, H. K., Green, P. D., McIntyre, M. D., & James, L. R. (2007). An integrative typology of personality assessment for aggression: Implications for predicting counterproductive workplace behavior. *Journal of Applied Psychology*, 92(3), 722–744. <https://doi.org/10.1037/0021-9010.92.3.722>
- Bolino, M. C., Harvey, J., & Bachrach, D. G. (2012). A self-regulation approach to understanding citizenship behavior in organizations. *Organizational Behavior and Human Decision Processes*, 119(1), 126–139. <https://doi.org/10.1016/j.obhdp.2012.05.006>
- Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, 10(2), 99–109. https://doi.org/10.1207/s15327043hup1002_3
- Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, 79(4), 631–643. <https://doi.org/10.1037/0022-3514.79.4.631>
- Bourdage, J. S., Lee, K., Lee, J. H., & Shin, K. H. (2012). Motives for organizational citizenship behavior: Personality correlates and coworker ratings of OCB. *Human Performance*, 25(3), 179–200. <https://doi.org/10.1080/08959285.2012.683904>
- Boyatzis, R. E., Goleman, D., & Rhee, K. S. (2000). Clustering competence in emotional intelligence. In R. Bar-On & J. D. A. Parker (Eds.), *The handbook of emotional intelligence* (pp. 343–362). Jossey-Bass.
- Byrne, B. (1998). *Structural equation modeling*. Erlbaum.
- Camara, W. J., & Schneider, D. L. (1994). Integrity tests: Facts and unresolved issues. *American Psychologist*, 49(2), 112–119. <https://doi.org/10.1037/0003-066x.49.2.112>
- Chuah, S. C., Drasgow, F., & Roberts, B. W. (2006). Personality assessment: Does the medium matter? no. *Journal of Research in Personality*, 40(4), 359–376. <https://doi.org/10.1016/j.jrp.2005.01.006>
- Church, A. T., & Katigbak, S. K. (2017). Trait consistency and validity across cultures: Examining trait and cultural psychology perspectives. In A. T. Church (Ed.), *The Praeger handbook of personality across cultures* (pp. 279–308). Praeger.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Colquitt, J. A., Sabey, T. B., Rodell, J. B., & Hill, E. T. (2019). Content validation guidelines: Evaluation criteria for definitional correspondence and definitional distinctiveness. *Journal of Applied Psychology*, 104(10), 1243–1265. <https://doi.org/10.1037/apl0000406>

- Colquitt, J. A., Scott, B. A., Rodell, J. B., Long, D. M., Zapata, C. P., Conlon, D. E., & Wesson, M. J. (2013). Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *The Journal of Applied Psychology*, 98(2), 199–236. <https://doi.org/10.1037/a0031757>
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136(6), 1092–1122. <https://doi.org/10.1037/a0021212>
- Connelly, B. S., Ones, D. S., & Hülshager, U. R. (2018). Personality in industrial, work and organizational psychology: Theory, measurement and application. In D. S. Ones, N. Anderson, C. Viswesvaran, & H. K. Sinangil (Eds.), *The SAGE handbook of industrial, work and organizational psychology* (2nd ed., Vol. 1, pp. 320–365). Sage. <https://doi.org/10.4135/9781473914940.n13>
- Cseh, M., Davis E, B., & Khilji S, E. (2013). Developing a global mindset: Learning of global leaders. *European Journal of Training & Development*, 37(5), 489–499. <https://doi.org/10.1108/03090591311327303>
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, 12(2), 163–170. <https://doi.org/10.1111/1467-9280.00328>
- Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of student's t-test. *International Review of Social Psychology*, 30(1), 92–101. <https://doi.org/10.5334/irsp.82>
- De Raad, B., Barelds, D. P. H., Timmerman, M. E., De Roover, 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>
- DeSimone, J. A., & James, L. R. (2015). An item analysis of the conditional reasoning test of aggression. *Journal of Applied Psychology*, 100(6), 1872–1886. <https://doi.org/10.1037/apl0000026>
- De Vries, R. E. (2012). Personality predictors of leadership styles and the self–other agreement problem. *The Leadership Quarterly*, 23(5), 809–821. <https://doi.org/10.1016/j.leaqua.2012.03.002>
- De Vries, R. E. (2013). The 24-item brief HEXACO inventory (BHI). *Journal of Research in Personality*, 47(6), 871–880. <https://doi.org/10.1016/j.jrp.2013.09.003>
- De Vries, R. E., Ashton, M. C., & Lee, K. (2009). De zes belangrijkste persoonlijkheidsdimensies en de HEXACO Persoonlijkheidsvragenlijst [The six most important personality dimensions and the HEXACO Personality Inventory]. *Gedrag & Organisatie*, 22(3), 208–250. <https://doi.org/10.5117/2009.022.003.004>
- De Vries, R. E., Pathak, R. D., Van Gelder, J. L., & Singh, G. (2017). Explaining unethical business decisions: The role of personality, environment, and states. *Personality and Individual Differences*, 117, 188–197. <https://doi.org/10.1016/j.paid.2017.06.007>
- De Vries, R. E., & Van Gelder, J.-L. (2015). Explaining workplace delinquency: The role of honesty–humility, ethical culture, and employee surveillance. *Personality and Individual Differences*, 86, 112–116. <https://doi.org/10.1016/j.paid.2015.06.008>
- Dietl, E., & Meurs, J. A. (2019). Implicit core self-evaluations and work outcomes: Validating an indirect measure. *Journal of Occupational and Organizational Psychology*, 92(1), 169–190. <https://doi.org/10.1111/joop.12244>
- Dijksterhuis, A., & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science*, 1(2), 95–109. <https://doi.org/10.1111/j.1745-6916.2006.00007.x>
- Dunning, D. (2012). The relation of self to social perception. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (2nd ed., pp. 481–501). Guilford Press.
- Dunning, D., & Cohen, G. L. (1992). Egocentric definitions of traits and abilities in social judgment. *Journal of Personality and Social Psychology*, 63(3), 341–355. <https://doi.org/10.1037/0022-3514.63.3.341>
- Dunning, D., & Hayes, A. F. (1996). Evidence for egocentric comparison in social judgment. *Journal of Personality and Social Psychology*, 71(2), 213–229. <https://doi.org/10.1037/0022-3514.71.2.213>
- Egloff, B., Schwerdtfeger, A., & Schmukle, S. C. (2005). Temporal stability of the implicit association test-anxiety. *Journal of Personality Assessment*, 84(1), 82–88. https://doi.org/10.1207/s15327752jpa8401_14
- Eidelman, S., & Biernat, M. (2007). Getting more from success: Standard raising as esteem maintenance. *Journal of Personality and Social Psychology*, 92(5), 759–774. <https://doi.org/10.1037/0022-3514.92.5.759>
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49(8), 709–724. <https://doi.org/10.1037/0003-066x.49.8.709>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology*, 23, 75–109. [https://doi.org/10.1016/S0065-2601\(08\)60318-4](https://doi.org/10.1016/S0065-2601(08)60318-4)
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54(1), 297–327. <https://doi.org/10.1146/annurev.psych.54.101601.145225>
- Fine, S. (2010). Cross-cultural integrity testing as a marker of regional corruption rates. *International Journal of Selection and Assessment*, 18(3), 251–259. <https://doi.org/10.1111/j.1468-2389.2010.00508.x>
- Funder, D. C. (2002). Personality psychology: Current status and some issues for the future. *Journal of Research in Personality*, 36(6), 638–639. [https://doi.org/10.1016/S0092-6566\(02\)00515-9](https://doi.org/10.1016/S0092-6566(02)00515-9)
- Galić, Z., Scherer, K. T., & LeBreton, J. M. (2014). Validity evidence for a Croatian version of the conditional reasoning test for aggression. *International Journal of Selection and Assessment*, 22(4), 343–354. <https://doi.org/10.1111/ijsa.12082>

- Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment*, 11(4), 340–344. <https://doi.org/10.1111/j.0965-075x.2003.00256.x>
- Goodman, S. A., & Svyantek, D. J. (1999). Person–organization fit and contextual performance: Do shared values matter. *Journal of Vocational Behavior*, 55(2), 254–275. <https://doi.org/10.1006/jvbe.1998.1682>
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4–27. <https://doi.org/10.1037/0033-295X.102.1.4>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, 36(3), 341–355. <https://doi.org/10.1108/00483480710731310>
- Gustafsson, J.-E., & Åberg-Bengtsson, L. (2010). Unidimensionality and interpretability of psychological instruments. In S. E. Embretson (Ed.), *Measuring psychological constructs: Advances in model-based approaches* (pp. 97–121). American Psychological Association. <https://doi.org/10.1037/12074-005>
- Harari, M., & Viswesvaran, C. (2018). Individual job performance. In D. S. Ones, N. Anderson, & C. Viswesvaran (Eds.), *The sage handbook of industrial, work and organizational psychology* (pp. 55–72). SAGE Publications Ltd.
- Harrington, R., & Loffredo, D. A. (2011). Insight, rumination, and self-reflection as predictors of well-being. *The Journal of Psychology*, 145(1), 39–57. <https://doi.org/10.1080/00223980.2010.528072>
- He, J., Buchholz, J., & Klieme, E. (2017). Effects of anchoring vignettes on comparability and predictive validity of student self-reports in 64 cultures. *Journal of Cross-Cultural Psychology*, 48(3), 319–334. <https://doi.org/10.1177/0022022116687395>
- Hilbig, B. E., & Zettler, I. (2009). Pillars of cooperation: Honesty–humility, social value orientations, and economic behavior. *Journal of Research in Personality*, 43(3), 516–519. <https://doi.org/10.1016/j.jrp.2009.01.003>
- Hodson, G., Book, A., Visser, B. A., Volk, A. A., Ashton, M. C., & Lee, K. (2018). Is the dark triad common factor distinct from low honesty–humility? *Journal of Research in Personality*, 73, 123–129. <https://doi.org/10.1016/j.jrp.2017.11.012>
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31(10), 1369–1385. <https://doi.org/10.1177/0146167205275613>
- Hopkins, D. J., & King, G. (2010). Improving anchoring vignettes: Designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly*, 74(2), 201–222. <https://doi.org/10.1093/poq/nfq011>
- Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist*. Sage. <https://doi.org/10.4135/9780857028075>
- James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods*, 1(2), 131–163. <https://doi.org/10.1177/109442819812001>
- James, L. R., & LeBreton, J. M. (2012). *Assessing the implicit personality through conditional reasoning*. American Psychological Association. <https://doi.org/10.1037/13095-000>
- Jaramillo, F., Carrillat, F. A., & Locander, W. B. (2005). A meta-analytic comparison of managerial ratings and self-evaluations. *Journal of Personal Selling & Sales Management*, 25(4), 315–328. <https://doi.org/10.1080/08853134.2005.10749067>
- Johnson, M. K., Rowatt, W. C., & Petrini, L. (2011). A new trait on the market: Honesty–humility as a unique predictor of job performance ratings. *Personality and Individual Differences*, 50(6), 857–862. <https://doi.org/10.1016/j.paid.2011.01.011>
- Julan, A. M., Novitsky, C., Lee, K., & Ashton, M. C. (2022). Convergent validity of three brief six-factor measures of personality. *Personality and Individual Differences*, 188, 111436. <https://doi.org/10.1016/j.paid.2021.111436>
- Kacmar, K. M., Witt, L. A., Zivnuska, S., & Gully, S. M. (2003). The interactive effect of leader-member exchange and communication frequency on performance ratings. *Journal of Applied Psychology*, 88(4), 764–772. <https://doi.org/10.1037/0021-9010.88.4.764>
- Kaiser, H. F. (1970). A second-generation little jiffy. *Psychometrika*, 35(4), 401–415. <https://doi.org/10.1007/bf02291817>
- Kasten, N., & Freund, P. A. (2015). A meta-analytical multilevel reliability generalization of situational judgment tests (SJTs). *European Journal of Psychological Assessment*, 32(3), 230–240. <https://doi.org/10.1027/1015-5759/a000250>
- Kelloway, E. K., Loughlin, C., Barling, J., & Nault, A. (2002). Self-reported counterproductive behaviors and organizational citizenship behaviors: Separate but related constructs. *International Journal of Selection and Assessment*, 10(1–2), 143–151. <https://doi.org/10.1111/1468-2389.00201>
- King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *The American Political Science Review*, 98(1), 191–207. <https://doi.org/10.1017/s000305540400108x>
- King, G., & Wand, J. (2007). Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes. *Political Analysis*, 15(1), 46–66. <https://doi.org/10.1093/pan/aml011>
- Kirkman, B. L., & Shapiro, D. L. (2001). The impact of team members' cultural values on productivity, cooperation, and empowerment in self-managing work teams. *Journal of Cross-Cultural Psychology*, 32(5), 597–617. <https://doi.org/10.1177/0022022101032005005>

- Lane, S., Raymond, M. R., Haladyna, T. M., & Downing, S. M. (2016). Test development process. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of test development* (2nd ed., pp. 3–18). Routledge. <https://doi.org/10.4324/9780203102961>
- Lang, J. W., Zettler, I., Ewen, C., & Hülshager, U. R. (2012). Implicit motives, explicit traits, and task and contextual performance at work. *Journal of Applied Psychology*, 97(6), 1201–1217. <https://doi.org/10.1037/a0029556>
- Leavitt, K., Fong, C. T., & Greenwald, A. G. (2011). Asking about well-being gets you half an answer: Intra-individual processes of implicit and explicit job attitudes. *Journal of Organizational Behavior*, 32(4), 672–687. <https://doi.org/10.1002/job.746>
- LeBel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin*, 37(4), 570–583. <https://doi.org/10.1177/0146167211400619>
- LeBreton, J. M., Barksdale, C. D., Robin, J., & James, L. R. (2007). Measurement issues associated with conditional reasoning tests: Indirect measurement and test faking. *Journal of Applied Psychology*, 92(1), 1–16. <https://doi.org/10.1037/0021-9010.92.1.1>
- Lee, K., & Allen, N. J. (2002). Organizational citizenship behavior and workplace deviance: The role of affect and cognitions. *Journal of Applied Psychology*, 87(1), 131–142. <https://doi.org/10.1037/0021-9010.87.1.131>
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39(2), 329–358. https://doi.org/10.1207/s15327906mbr3902_8
- Lee, K., & Ashton, M. C. (2020). Sex differences in HEXACO personality characteristics across countries and ethnicities. *Journal of Personality*, 88(6), 1075–1090. <https://doi.org/10.1111/jopy.12551>
- Lee, K., Ashton, M. C., & De Vries, R. E. (2005). Predicting workplace delinquency and integrity with the HEXACO and five-factor models of personality structure. *Human Performance*, 18(2), 179–197. https://doi.org/10.1207/s15327043hup1802_4
- Lee, Y., Berry, C. M., & Gonzalez-Mulé, E. (2019). The importance of being humble: A meta-analysis and incremental validity analysis of the relationship between honesty-humility and job performance. *Journal of Applied Psychology*, 104(12), 1535–1546. <https://doi.org/10.1037/apl0000421>
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1(2), 27–66. <https://doi.org/10.1111/1529-1006.002>
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149. <https://doi.org/10.1037/1082-989X.1.2.130>
- Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big five, or one in addition? *Personnel Psychology*, 60(1), 1–34. <https://doi.org/10.1111/j.1744-6570.2007.00063.x>
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224–253. <https://doi.org/10.1037/0033-295x.98.2.224>
- McAllister, D. J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1), 24–59. <https://doi.org/10.2307/256727>
- McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant perspectives during selection: A review addressing “so what?,” “What’s new?,” and “where to next?”. *Journal of Management*, 43(6), 1693–1725. <https://doi.org/10.1177/0149206316681846>
- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, 15(1), 28–50. <https://doi.org/10.1177/1088868310366253>
- McDowall, A., & Redman, A. (2017, April 5). *Psychological assessment – an overview of theoretical, practical and industry trends* [Video]. YouTube. <https://youtu.be/Sa-kU5qwile>
- Meier, K. J., & O’Toole, L. J., Jr. (2013). I think (I am doing well), therefore I am: Assessing the validity of administrators’ self-assessments of performance. *International Public Management Journal*, 16(1), 1–27. <https://doi.org/10.1080/10967494.2013.796253>
- Mitchell, M. S., & Ambrose, M. L. (2007). Abusive supervision and workplace deviance and the moderating effects of negative reciprocity beliefs. *Journal of Applied Psychology*, 92(4), 1159–1168. <https://doi.org/10.1037/0021-9010.92.4.1159>
- Moors, A., Spruyt, A., & De Houwer, J. (2010). In search of a measure that qualifies as implicit: Recommendations based on a decomposition view of automaticity. In B. Gawronski & K. B. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 19–37). Guilford Press.
- Morgan, C. D., & Murray, H. A. (1935). A method for investigating fantasies: The thematic apperception test. *Archives of Neurology & Psychiatry*, 34(2), 289–306. <https://doi.org/10.1001/archneurpsyc.1935.02250200049005>
- O’Boyle, E. H., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2012). A meta-analysis of the dark triad and work behavior: A social exchange perspective. *Journal of Applied Psychology*, 97(3), 557–579. <https://doi.org/10.1037/a0025679>
- Oh, I.-S., Lee, K., Ashton, M. C., & De Vries, R. E. (2011). Are dishonest extraverts more harmful than dishonest introverts? The interaction effects of honesty-humility and extraversion in predicting workplace deviance. *Applied Psychology: An International Review*, 60(3), 496–516. <https://doi.org/10.1111/j.1464-0597.2011.00445.x>

- Oh, I. S., Le, H., Whitman, D. S., Kim, K., Yoo, T. Y., Hwang, J. O., & Kim, C. S. (2014). The incremental validity of honesty-humility over cognitive ability and the big five personality traits. *Human Performance*, 27(3), 206–224. <https://doi.org/10.1080/08959285.2014.913594>
- 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>
- Oostrom, J. K., Köbis, N. C., Ronay, R., & Cremers, M. (2017). False consensus in situational judgment tests: What would others do? *Journal of Research in Personality*, 71, 33–45. <https://doi.org/10.1016/j.jrp.2017.09.001>
- Organ, D. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington Books.
- Passmore, J. (2012). *Psychometrics in coaching: Using psychological and psychometric tools for development* (2nd Ed.). Kogan Page.
- Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556–563. [https://doi.org/10.1016/s0092-6566\(02\)00505-6](https://doi.org/10.1016/s0092-6566(02)00505-6)
- Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, 94(1), 16–31. <https://doi.org/10.1037/0022-3514.94.1.16>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>
- 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>
- Pletzer, J. L., Oostrom, J. K., Bentvelzen, M., & De Vries, R. E. (2020). Comparing domain-and facet-level relations of the HEXACO personality model with workplace deviance: A meta-analysis. *Personality and Individual Differences*, 152, Article 109539. <https://doi.org/10.1016/j.paid.2019.109539>
- Pletzer, J. L., Oostrom, J. K., & De Vries, R. E. (2021). HEXACO personality and organizational citizenship behavior: A domain-and facet-level meta-analysis. *Human Performance*, 2020(1), 1–22. <https://doi.org/10.1080/08959285.2021.1891072>
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61(1), 153–172. <https://doi.org/10.1111/j.1744-6570.2008.00109.x>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Primi, R., Zanon, C., Santos, D., De Fruyt, F., & John, O. P. (2016). Anchoring vignettes: Can they make adolescent self-reports of social-emotional skills more reliable, discriminant, and criterion-valid? *European Journal of Psychological Assessment*, 32(1), 39–51. <https://doi.org/10.1027/1015-5759/a000336>
- Protzko, J., & Schooler, J. W. (2019). Kids these days: Why the youth of today seem lacking. *Science Advances*, 5(10), Article eaav5916. <https://doi.org/10.1126/sciadv.aav5916>
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544–559. <https://doi.org/10.1080/00223891.2010.496477>
- Ridgeway, C. (2017). Projective measures and occupational assessment. In B. Cripps (Ed.), *Psychometric testing: Critical perspectives* (pp. 213–220). John Wiley & Sons. <https://doi.org/10.1002/9781119183020>
- Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal*, 38(2), 555–572. <https://doi.org/10.2307/256693>
- Robles, M. M. (2012). Executive perceptions of the top 10 soft skills needed in today's workplace. *Business Communication Quarterly*, 75(4), 453–465. <https://doi.org/10.1177/1080569912460400>
- Rogelberg, S. G., & Stanton, J. M. (2007). Introduction: Understanding and dealing with organizational survey nonresponse. *Organizational Research Methods*, 10(2), 195–209. <https://doi.org/10.1177/1094428106294693>
- 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>
- Rudolph, A., Schröder-Abé, M., Schütz, A., Gregg, A. P., & Sedikides, C. (2008). Through a glass, less darkly? Reassessing convergent and discriminant validity in measures of implicit self-esteem. *European Journal of Psychological Assessment*, 24(4), 273–281. <https://doi.org/10.1027/1015-5759.24.4.273>
- Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, 26(3), 565–606. <https://doi.org/10.1177/014920630002600308>
- Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology*, 102(3), 254–273. <https://doi.org/10.1037/apl0000151>
- Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education. *American Psychologist*, 56(4), 302–318. <https://doi.org/10.1037/0003-066x.56.4.302>
- 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>

- Schultheiss, O. C., Liening, S. H., & Schad, D. (2008). The reliability of a picture story exercise measure of implicit motives: Estimates of internal consistency, retest reliability, and ipsative stability. *Journal of Research in Personality*, 42(6), 1560–1571. <https://doi.org/10.1016/j.jrp.2008.07.008>
- Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. Yale University Press.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107–120. <https://doi.org/10.1007/s11336-008-9101-0>
- Smith, C. A., Organ, D. W., & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology*, 68(4), 653–663. <https://doi.org/10.1037/0021-9010.68.4.653>
- Spector, P. E., Bauer, J. A., & Fox, S. (2010). Measurement artifacts in the assessment of counterproductive work behavior and organizational citizenship behavior: Do we know what we think we know? *Journal of Applied Psychology*, 95(4), 781–790. <https://doi.org/10.1037/a0019477>
- Steffens, M. C. (2004). Is the implicit association test immune to faking? *Experimental Psychology*, 51(3), 165–179. <https://doi.org/10.1027/1618-3169.51.3.165>
- Thielmann, I., Moshagen, M., Hilbig, B., & Zettler, I. (2021). On the comparability of basic personality models: Meta-analytic correspondence, scope, and orthogonality of the big five and HEXACO dimensions. *European Journal of Personality*, 36(6), 870–900. <https://doi.org/10.1177/08902070211026793>
- Udayar, S., Fiori, M., Thalmayer, A. G., & Rossier, J. (2018). Investigating the link between trait emotional intelligence, career indecision, and self-perceived employability: The role of career adaptability. *Personality and Individual Differences*, 135, 7–12. <https://doi.org/10.1016/j.paid.2018.06.046>
- Uhlmann, E. L., Leavitt, K., Menges, J. I., Koopman, J., Howe, M., & Johnson, R. E. (2012). Getting explicit about the implicit. *Organizational Research Methods*, 15(4), 553–601. <https://doi.org/10.1177/1094428112442750>
- Van Rensburg, J. Y., De Kock, F. S., & Deros, E. (2019). 'Going implicit': Using implicit measures in organizations. *Gedrag en Organisatie*, 32(3), 131–161.
- Van Rensburg, Y. E. J., De Kock, F., De Vries, R. E., & Deros, E. (2022). Measuring honesty-humility with an implicit association test (IAT): Construct and criterion validity. *Journal of Research in Personality* 99, 104234. Article 104234. <https://doi.org/10.1016/j.jrp.2022.104234>
- Vardi, Y., & Weitz, E. (2004). *Misbehavior in organizations: Theory, research, and management*. Lawrence Erlbaum Associates Publishers.
- Vargas, P., Sekaquaptewa, D., & von Hippel, W. (2007). Armed only with paper and pencil: "low-tech" measures of implicit attitudes. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 103–124). Guilford Press.
- Vargas, P. T., von Hippel, W., & Petty, R. E. (2004). Using partially structured attitude measures to enhance the attitude-behavior relationship. *Personality and Social Psychology Bulletin*, 30(2), 197–211. <https://doi.org/10.1177/0146167203259931>
- Vazire, S., & Carlson, E. N. (2010). Self-knowledge of personality: Do people know themselves? *Social and Personality Psychology Compass*, 4(8), 605–620. <https://doi.org/10.1111/j.1751-9004.2010.00280.x>
- Vecchione, M., Dentale, F., Alessandri, G., & Barbaranelli, C. (2014). Fakability of implicit and explicit measures of the big five: Research findings from organizational settings. *International Journal of Selection and Assessment*, 22(2), 211–218. <https://doi.org/10.1111/ijsa.12070>
- Wiita, N. E., Meyer, R. D., Kelly, E. D., & Collins, B. J. (2020). Not aggressive or just faking it? Examining faking and faking detection on the conditional reasoning test of aggression. *Organizational Research Methods*, 23(1), 96–123. <https://doi.org/10.1177/1094428117703685>
- Wright, N. A., & Meade, A. W. (2011, April 13–16). *Predictive validity and procedural justice of the implicit association test*. 26th Annual Conference Program of the APA Division 14 Society for Industrial and Organizational Psychology (SIOP), Chicago, IL.
- Ziegler, M., Kemper, C. J., & Kruey, P. (2014). Short scales – five misunderstandings and ways to overcome them. *Journal of Individual Differences*, 35(4), 185–189. <https://doi.org/10.1027/1614-0001/a000148>