



The 24-item Brief HEXACO Inventory (BHI)

Reinout E. de Vries

VU University Amsterdam, Department of Social and Organizational Psychology, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands



ARTICLE INFO

Article history:

Available online 20 September 2013

Keywords:

HEXACO

Short measures

Test construction

Reliability

Validity

Generalizability theory

ABSTRACT

Up until now, no really short instrument that measures the six personality dimensions of the HEXACO model has been available. In two studies, I report the construction of the Brief HEXACO Inventory (BHI), which represents the 24 HEXACO facets with 1 item per facet (i.e., 4 items per domain) and which takes approximately 2–3 min to complete. Although characterized by relatively low alpha reliability, its test–retest stability, self–other agreement, and convergent correlations with full-length scales are relatively high and its validity loss is only modest. Correcting for attenuation using a weighted average of alpha reliability, test–retest stability, and self–other agreement, the BHI re-estimates the original construct validity correlations of the HEXACO-PI-R with relatively great accuracy.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Researchers often shy away from short personality scales because of their low alpha reliabilities. Recently, however, evidence has amassed that the effects on validity of well-constructed short personality scales with low alpha reliabilities are not as large as previously feared. As a consequence, in the past decade a great number of short personality instruments have become available (e.g., [Credé, Harms, Niehorster, & Gaye-Valentine, 2012](#)). Almost all of the short personality measures constructed so far pertain to the Big Five model, whereas lexical research has provided evidence for six – instead of five – main personality dimensions ([Ashton et al., 2004](#)). These six are referred to as the HEXACO dimensions of personality ([Lee & Ashton, 2004](#)). In this study, I will report the development of the Brief HEXACO Inventory (BHI), a 24-item instrument that is meant to cover these main six dimensions of personality with relatively low loss of validity.

1.1. Advantage and (purported) disadvantages of short personality scales

Short personality scales are defined in this study as containing anywhere between 1 and 5 items per domain scale. The advantages and disadvantages of short personality scales are known ([Credé et al., 2012](#)), but sometimes not adequately interpreted. The main advantage of short personality scales is that they require minimal time investment of respondents. Short personality scales are especially useful when respondents are asked to fill out lengthy questionnaires with lots of other constructs in large-scale surveys.

They are also useful when the personality instrument is part of an Experience Sampling Method (ESM) study, in which respondents only have little time to fill out a questionnaire. In large-scale and ESM studies, including a personality instrument with 50 or more questions is often not an option. The choice is often between including a short personality instrument or not including a personality instrument at all. When faced with such forced-choice dilemmas, most researchers are likely to prefer including a short instrument, rather than not including anything at all.

The main purported disadvantages of short personality scales are (1) low reliability, (2) low validity, and (3) low (diagnostic) utility. Especially the importance of alpha reliability ([Cronbach, 1951](#)) is often exaggerated (e.g., [McCrae, Kurtz, Yamagata, & Terracciano, 2011](#); [Sijtsma, 2009](#)), causing researchers to avoid constructing or using short personality scales. When researchers do decide to construct a short personality scale, they often struggle with the problem of a trade-off between coverage and consistency. They have to determine whether to optimize reliability and thus to narrow the scope of the personality scale or to optimize coverage of the construct at the expense of its internal consistency.¹ To give an example based on items from the HEXACO-PI-R ([Ashton & Lee, 2008](#); [Lee & Ashton, 2004](#)), in a large-scale Dutch community sample with $N = 1352$ ([De Vries, Ashton, & Lee, 2009](#)), items from the same

¹ Please note that internal consistency and the most often used reliability estimate, [Cronbach's \(1951\)](#) alpha reliability coefficient, refer to two different things. Alpha reliability of a test depends on the average intercorrelation of its items and the test length whereas internal consistency refers to the strength of the intercorrelations of its items and the degree to which the items are unidimensional. A test that contains items that show a multidimensional structure may still be highly reliable according to the alpha reliability coefficient when the test has sufficient length, but should not be referred to as 'internally consistent'. See [Sijtsma \(2009\)](#) for an overview.

E-mail address: re.de.vries@vu.nl

personality facet correlate on average .30, but items from different facets within the same broad personality domain correlate on average .16. Based on the Spearman–Brown prophecy formula, scales based on items from the same facet will have the following expected alpha reliabilities with respectively 1, 2, 3, 4, or 5 items: .30, .46, .56, .63, and .68. In comparison, scales based on items from different facets will have the following substantially lower expected alpha reliabilities: .16, .28, .36, .43, and .49.² Whereas three items from the same facet would be enough to obtain an alpha reliability >.55, even six items would not be enough to obtain a similar result when using items from different facets of the same domain.

Researchers may opt to tackle this problem by choosing consistency over coverage, thus increasing alpha reliability, but introducing tautology problems. That is, items that are thought to represent a ‘central facet’ of the broad personality trait, and which correlate as strongly as possible with the broad personality domain and with each other, are likely to be tautological in content, producing ‘bloated specifics’ as scales (Cattell, 1973). Items which are tautological are likely to be susceptible to systematic response biases such as ‘transient errors’ (Becker, 2000; Chmielewski & Watson, 2009; Schmidt, Le, & Ilies, 2003; Thorndike, 1951), which refer to variations of responses across time due to fluctuating psychological states of the participant. Transient errors tend to produce consistent responses in the same response session but inconsistent responses across sessions. As such, transient errors are likely to increase alpha reliabilities, but to decrease test–retest stability estimates, self-other agreement correlations, and validity estimates.

Although reliability is most often associated with *internal* consistency (but: see footnote 1), *temporal* consistency (e.g., stability or test–retest reliability) and *rater* consistency (e.g., self-other agreement and/or inter-rater agreement) may sometimes provide a better estimate of a measure’s ‘true’ (unbiased) systematic variance.³ Temporal consistency can be ascertained by measuring the same construct a few weeks or months later, thus minimizing the chance of systematic personality changes while also minimizing spill-over memory effects. Research shows that internal consistency and temporal consistency may be only marginally related (e.g., $r = .25$ in Chmielewski and Watson’s (2009) study). Although internal consistency may be attenuated when choosing for coverage over consistency in the coverage – consistency trade-off, temporal consistency may actually suffer less or not at all from such an approach. Thus, scales based on items from different facets of a personality domain, which may show relatively low levels of internal consistency (e.g., alpha reliability), may actually show similar or even higher levels of temporal consistency (e.g., test–retest reliability) than scales based on items selected from similar facets.

The same is true for rater consistency (e.g., self-other agreement). Self-ratings with inflated alpha reliability estimates due to systematic response biases will not show higher self-other agreement correlations than self-ratings without such biases and, consequently, with lower alpha reliabilities. Rater consistency estimates do not suffer from the same biases that may affect temporal consistency estimates. For instance, (unconscious) self-deceptive self-enhancement (Paulhus, Bruce, & Trapnell, 1995) is likely to inadvertently increase temporal consistency but not rater consistency. However, rater consistency estimates do suffer from

numerous other biases associated with the personality judgment process (e.g., low levels of acquaintanceship; Funder, 1995), and thus self-other agreement estimates are generally lower than (test–retest) stability estimates. Furthermore, evidence of temporal and rater consistency is not a panacea for problems associated with internal consistency, because temporal and rater consistency can also exist in measures that consist of items that are unrelated. To give an extreme example, one- or two-months temporal consistency and even (well-acquainted) rater consistency may be very high of a measure that consists of items such as a person’s shoe size, the last two digits of his/her telephone number, and his/her residence number, but such a measure does not contain any true or interpretable systematic variance. That is, a scale that is characterized by a high level of heterogeneity may still show substantial temporal and rater consistency as long as each of the items in the scale shows high levels of test–retest stability and self-other agreement.⁴

The above makes clear that reliability estimates, whether measured using alpha coefficients, temporal consistency coefficients, or rater consistency coefficients, are seldom sufficient to evaluate the psychometric adequacy of a scale. For a personality scale, the proof of its adequacy for research – and practical – purposes is in its validity. Everything else remaining equal, the (correction for) attenuation formula teaches us that for short scales a reduction in reliability results in a reduction in validity which is equal to the square root of the fraction by which its reliability is reduced. When reliability estimates are compromised by systematic response biases such as transient errors, a reduction in reliability may be accompanied by an even less severe reduction in validity as long as a substantial part of the reduction in reliability is due to a reduction in its systematic response bias. Consequently, although shortening scales will tend to lower validity, short scales can optimize validity by reducing as much as possible potential response biases and by retaining as much as possible the underlying true systematic variance.

Last but not least, with lower reliability, utility suffers from shortening scales because both the estimated true score and the confidence interval around the true score are a function of a scale’s reliability (Nunnally, 1981). Therefore, scales with reliabilities substantially lower than .90 are deemed to be generally unfit for diagnostic purposes in assessment situations. Consequently, the main use for scales with low levels of reliability but still significant levels of ‘true’ systematic variance lies in exploratory research in large representative samples, which, after correction for attenuation, offer an opportunity to estimate the levels of association between personality and other variables.

1.2. Short Big Five instruments

There are currently a number of short domain-level personality scales – consisting of 1–5 items per scale – available, most notable the Ten-Item Personality Inventory (TIPI, Gosling, Rentfrow, Swann, 2003), the abridged Big Five (Langford, 2003), the Single-Item Measures of Personality (SIMP, Woods & Hampson, 2005), the Domain Self Ratings scale (DSRs, Bernard, Walsh, & Mills, 2005), the 20-item Mini International Personality Item Pool personality measure (Mini-IPIP, Donnellan, Oswald, Baird, & Lucas, 2006), the Five-Item Measure of the Big Five (Aronson, Reilly, & Lynn, 2006), and the Big Five Inventory-10 (BFI-10, Rammstedt & John, 2007). All of the above scales purport to measure the five

² Note that higher alpha reliabilities can be obtained when selecting items that show the highest levels of corrected item–total correlations. Consequently, in practice, short scales that are based on a selection which includes the best performing items will have higher alpha reliabilities (Paunonen & Jackson, 1985).

³ McCrae et al. (2011) use rater consistency – or cross-observer agreement – as indication of ‘consensual validity’ instead of reliability. Because raters in a design with multiple well-acquainted raters provide systematic variance, unique variance, and error variance on the personality variables of interest, in line with Hofstee (1994) and John and Soto (2007), it can be argued that it makes more sense to treat inter-rater agreement as a proxy of reliability instead of validity.

⁴ The same is true for heritability estimates; scales that show high levels of heterogeneity may still show substantial levels of heritability. This is probably one of the reasons why heritability estimates of the purported General Factor of Personality (GFP) are highly similar to heritability estimates of domain-level personality scales (e.g., Rushton et al., 2009).

domains of the Big Five/Five Factor Model. However, they differ with respect to the measurement method. Some use items that consist of multiple adjectives (e.g., Aronson et al., 2006; Gosling et al., 2003; Langford, 2003) whereas other scales consist of simple, unifacted sentences (Donnellan et al., 2006; Rammstedt & John, 2007) or complex, multifaceted sentences (Bernard et al., 2005; Woods & Hampson, 2005); some use Likert-type answering scales (Aronson et al., 2006; Donnellan et al., 2006; Gosling et al., 2003; Rammstedt & John, 2007) whereas others use bipolar or multicategorical anchors (Bernard et al., 2005; Langford, 2003; Woods & Hampson, 2005). Note that the use of multiple adjectives, complex sentences, and bipolar or multicategorical anchors tend to make items more complex, potentially increasing response time – a shortening in response time being one of the most important reasons for employing short scales.

Although especially alpha reliability, but to a lesser extent temporal consistency and rater consistency have been found to be attenuated due to scale shortening, on the whole the above-mentioned investigators of short scales agree that validity does not seem to suffer to a great extent when using short scales (see also Burisch, 1997; Paunonen & Jackson, 1985). Although even one-item scales performed adequately in studies by Aronson et al. (2006), Bernard et al. (2005), and Woods and Hampson (2005), Credé et al. (2012) noted that this is probably the result of using more 'complex' single item scales, which have more extensively formulated response options but which, consequently, take more time to comprehend and to answer. Based on a comparison of a number of short scales, Credé et al. (2012) showed that substantial improvements in criterion validity are made when going from one to two items per scale and that moderate improvements are made when scales are further increased in length. Consequently, Credé et al. (2012) made the following four recommendations: (1) To avoid single item scales; (2) to add, if possible, a few extra items to single item scales because this will only marginally increase the time taken to fill out the questionnaire but will substantially decrease both Type 1 and Type 2 errors; (3) to avoid making strong claims about (incremental) validity when short scales are used, especially when using broad criteria; and (4) to use short scales mainly in research settings in which boredom, fatigue, or lack of interest may increase careless or random answering.

1.3. The HEXACO model

Until now, short broad-bandwidth personality scales have only been developed for the Big Five/FFM. However, recent evidence has suggested that the Big Five model needs to be revised to include an additional dimension, called Honesty–Humility, and to rearrange facets associated with two of its dimensions, i.e., Agreeableness and Emotional Stability (Ashton et al., 2004; Ashton & Lee, 2007). Based on lexical personality findings, Lee and Ashton (2004) have constructed the HEXACO Personality Inventory, a questionnaire operationalization of the HEXACO model, which consists of the following six broad domain scales: Honesty–Humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, and Openness to experience. Each of the domain scales in the HEXACO model is represented by four facets which capture related – but empirically distinguishable – behaviors. Due to its addition of Honesty–Humility dimension and its facets, the HEXACO model has been able to outperform the Big Five and/or FFM model in the explanation of several criteria associated with anti-social or self-serving behaviors, such as work-place and academic delinquency, criminal choices, sexual harassment, egoism, narcissism, status-driven risk-taking, Machiavellianism, and psychopathy (Ashton & Lee, 2008; Ashton, Lee, Pozzobon, Visser, & Worth, 2010; De Vries & Van Kampen, 2010; De Vries, De Vries, & Born, 2011; De Vries, De Vries, De

Hoogh, & Feij, 2009; Lee, Gizzarone, & Ashton, 2003; Lee et al., 2013; Van Gelder & De Vries, 2012).

Apart from anti-social and self-serving behaviors, Honesty–Humility has also been found to be an important predictor of values and socio-political attitudes (Hilbig & Zettler, 2009; Lee, Ashton, Ogunfowora, Bourdage, & Shin, 2010; Lee et al., 2009; Zettler & Hilbig, 2010). Both values and socio-political attitudes have been found to be described by two broadly independent dimensions, one referring to conservation (Conservation versus Openness to Change) and abiding to social conventions (Right-Wing Authoritarianism – RWA) and the other referring to preference for power or status (Self-Enhancement versus Self-Transcendence) and a hierarchical orientation (Social Dominance Orientation – SDO) (Duriez, Van Hiel, & Kossowska, 2005; Lee et al., 2010). Of the HEXACO personality dimensions, especially Openness to Experience and Honesty–Humility have been found to be relevant in the explanation of these two dimensions. Openness to Experience has been found to be (negatively) related to Social Conformity (e.g., Conservation and RWA) whereas Honesty–Humility has been found to be (negatively) related to Hierarchy Orientation (e.g., Self-Enhancement and SDO) (Lee et al., 2009, 2010). Openness to Experience in the HEXACO model is highly similar to Openness to Experience in the Big Five/FFM model, but, as noted above, the Big Five model lacks a dimension that sufficiently captures the variance associated with Honesty–Humility. Consequently, because of the apparent societal impact of Honesty–Humility in its relations with anti-social and self-serving behaviors and with socio-political values, a short scale that may be used in large-scale societal studies that incorporates Honesty–Humility seems to be especially warranted.

To summarize, although alpha reliability and utility are generally compromised when using a short personality scale, test–retest consistency, rater consistency, and validity seem to be generally less strongly affected. Unless the items are made very complex, 1-item scales are generally not preferred due to the negative effects on validity. Two-item scales offer a significant increment in validity, but may fail to capture the broad content of personality dimensions, especially when items are formulated tautological in order to optimize alpha reliability. To minimize transient errors and to optimize coverage, 3- or 4-item scales seem to offer the most optimal representation of personality constructs, while still keeping answering times as short as possible. Consequently, this study will report the construction of a short – 4 items per domain – personality questionnaire based on the HEXACO model. In two studies, consisting of a community sample and a student sample, I will report on the alpha reliability, test–retest reliability, self-other agreement, and validity of the 24-item Brief HEXACO Inventory, the BHI. Furthermore, I will present evidence on the amount of validity loss for different levels of correlations when comparing the original (200-item) HEXACO-PI-R with the BHI. In general, I will show that although alpha reliability is affected to a large extent, the effects on validity are relatively minor. Furthermore, I will show that the original HEXACO-PI-R correlations can be relatively straightforwardly and accurately estimated from the BHI correlations using a sample-weighted average of the alpha reliability, test–retest stability, and self-other agreement coefficients.

2. Method

2.1. Samples and procedures

Two separate studies were conducted to test the Brief HEXACO Inventory (BHI): a community study (Study 1) and a study among undergraduate psychology and educational science students (Study 2).

2.1.1. Study 1 – community sample

Several waves of data were collected using a large-scale internet panel. For this study, we used data from T1 (April 2008), T2 (October 2009), and T3 (November 2011). T1 consisted of three waves, waves 1, 2, and 3 ($N = 1,052$), which are more fully described in [De Vries and Van Kampen \(2010\)](#). From T2, which also consisted of three waves (waves 4–6), I used data from wave 4 ($N = 815$), which is described in [De Vries, Bakker-Pieper, Konings, and Schouten \(2013\)](#). From T3, which consisted of two waves (waves 7–8), I used wave 7 data. The final sample ($N = 525$) consisted of 273 (52.0%) women, with a mean age of 54.7 ($SD = 13.9$; range 22–87) years. Of the respondents, 147 (28.0%) had completed lower levels of education (e.g., primary education, lower-level secondary education, or lower-level tertiary education), 221 (42.1%) had completed medium levels of education (e.g., higher-level secondary education or medium-level tertiary education), and 157 (29.9%) had completed higher levels of education (e.g., higher-level tertiary education or university level education).

2.1.2. Study 2 – student sample

As part of two undergraduate courses, students provided self-ratings and obtained other-ratings on a number of different instruments in four waves (three waves in T1 and one wave in T2) of data collection during the first part of 2013. Valid data from undergraduates who provided permission to use the data for research purposes were obtained from 309 students (78.3% women; $M_{\text{age}} = 20.0$, $SD = 2.4$). From this sample, 307 well-acquainted other-ratings (62.2% women; $M_{\text{age}} = 24.5$, $SD = 10.9$) were obtained at T1 and 263 self-ratings (79.8% women; $M_{\text{age}} = 19.9$, $SD = 1.7$), were obtained two months later during another undergraduate course (T2, wave 4).

2.2. Instruments

2.2.1. HEXACO-PI-R

In both samples the full 200-item Dutch HEXACO-PI-R ([De Vries et al., 2009](#)) was used. The development of the HEXACO-PI-R, which is based on the main six lexical personality factors ([Ashton et al., 2004](#)) is more fully described in [Lee and Ashton \(2004\)](#). The HEXACO-PI-R measures each of the six personality domains, Honesty–Humility (H), Emotionality (E), eXtraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O), with 32 items (eight items for each of the four facets per domain scale). An additional eight items measure the interstitial facet Altruism ([Lee & Ashton, 2006](#)). All items were rated on a 1–5 (strongly disagree–strongly agree) scale. In previous research, the six domain scales have shown adequate psychometrics properties, with alpha reliabilities $>.80$ and generally low ($<.35$) domain scale intercorrelations ([De Vries et al., 2009](#); [Lee & Ashton, 2004](#)). The HEXACO-PI-R was obtained in the first wave in the community sample (Study 1) – 3.5 years prior to the BHI – and in the second wave of data collection in the student sample (Study 2), right before the BHI. The HEXACO-PI-R alpha reliabilities were .91/.92 for H, .89/.90 for E, .90/.90 for X, .89/.90 for A, .84/.90 for C, .87/.88 for O in Study 1 ($N = 525$)/Study 2 ($N = 309$) respectively.

2.2.2. Brief HEXACO Inventory (BHI)

The 24 item Brief HEXACO Inventory (BHI) is derived from another – recent – operationalization of the HEXACO model, the HEXACO Simplified Personality Inventory (HEXACO-SPI; [De Vries & Born, 2013](#)). In the HEXACO-SPI, new, short, and easy to comprehend items were written for all facets to make the inventory suitable for children and people with lower levels of (language) education. None of the HEXACO-SPI items, and thus none of the BHI items, are exactly the same as the HEXACO-PI-R items. In the process of constructing the BHI, I tried to optimize coverage and

convergent validity. Each of the six BHI domain scales is represented by four items. Coverage was optimized by choosing only one HEXACO-SPI item from each of the 24 facets (four facets per domain) and by making sure that one or three – and preferably two – of the items per domain were reverse keyed. Using the community sample (Study 1), convergent validity was optimized by choosing from each facet one BHI item from wave 7 with the highest correlation with the corresponding wave 1 HEXACO-PI-R domain scale. When items provided highly similar results, first of all I checked whether the items were sufficiently distinct from other domains and whether all domain items together covered the content of the domain adequately, and then I chose an item which optimized both the scale's convergent validity and its alpha reliability.

I compared this procedure (A) with a procedure that optimized alpha reliabilities from the start (B) and found that the convergent validities with the original HEXACO-PI-R of procedure A were generally higher than those of procedure B, although some of the alpha reliabilities were lower.⁵ The items were rated on a 1 (strongly disagree) to 5 (strongly agree) scale. A Principal Component Analysis (PCA) with varimax rotation was performed on the combined sample ($N = 834$). Six factors with eigenvalues > 1 were extracted. All items loaded $>.50$ on their designated factors (see [Supplemental Table S1](#)) with one exception. The 'Sincerity' item loaded equally (.34) on Honesty–Humility and Emotionality. In a separate PCA on the community sample, this item loaded .53 on the Honesty–Humility factor and did not have any significant cross-loadings, so this may be due to the different composition of the student sample. The overall descriptives and alpha reliabilities of Study 1 and Study 2 are reported in [Table 1](#). Note that in this Table, except for the alpha reliabilities and temporal consistency estimates, all Study 2 values refer to T1 – wave 2 data. For the sake of completeness, the alpha reliabilities of the other-ratings in Study 2 were .55 for H, .54 for E, .69 for X, .53 for A, .61 for C, and .60 for O, with a mean of .59. [Supplemental Table S2](#) provides additional correlational and descriptive information on both studies. To summarize Table S2, none of the BHI scales correlated $>.30$ with other BHI scales and the mean differences between men and women followed the pattern shown in other studies (e.g., [De Vries et al., 2009](#)). That is, compared to men, women scored mainly higher on Emotionality and Honesty–Humility. The self-assessment items of the original Dutch version used in this study and the translated and back-translated English version of the BHI are reported in Appendix A.

2.2.3. Other scales

To compare the validities of the HEXACO-PI-R and the BHI, in both Study 1 and Study 2 I selected 12 constructs which differed in strengths of relations with the HEXACO variables. From the community panel data used in Study 1, I selected the following instruments: the 5-Dimensional Personality Test (5DPT; [Van Kampen, 2012](#); obtained at T1 – wave 2), the Self-Report Psychopathy checklist version III (SRP-III; [Williams, Paulhus, & Hare, 2007](#); T1 – wave 3), and the Communication Styles Inventory (CSI; [De Vries et al., 2013](#); T2 – wave 4). The 5DPT consists of five psychopathological personality scales which are closely aligned to Five Factor Model scales ([Van Kampen, 2012](#)), each consisting of 20 dichotomous (yes–no) items. The alpha reliabilities of the 5DPT scales were .89 for Extraversion, .91 for Neuroticism, .87 for Absorption, .83 for Orderliness, and .84 for Insensitivity. The SRP-III consists of 20 items, rated on a 5-point (strongly disagree–strongly agree) scale, with an alpha reliability of .88 in this study. The CSI consists of 92 items, divided among six 5-point (strongly disagree–strongly

⁵ Further information about the selection procedure can be obtained from the author.

Table 1

BHI descriptives and reliabilities, and convergent/maximum discriminant correlations of the BHI with the HEXACO-PI-R.

Study (1 = Community/2 = Student)	Descriptives		Consistency					Convergent		Discriminant	
	<i>M(SD)</i>		Alpha			Temporal	Rater	<i>r</i> 's		Max. <i>r</i> 's	
	1	2	1	2 (T1)	2 (T2)	2	2	1	2	1	2
Honesty–Humility	3.85(.58)	3.77(.63)	.57	.55	.60	.78	.46	.59	.72	.18	.32
Emotionality	3.00(.57)	3.12(.70)	.46	.54	.56	.77	.58	.68	.81	–.21	–.28
eXtraversion	3.84(.59)	4.08(.52)	.72	.69	.63	.71	.45	.68	.78	.16	.18
Agreeableness	2.95(.53)	3.09(.57)	.44	.53	.49	.72	.39	.59	.73	.18	.28
Conscientiousness	3.51(.56)	3.10(.68)	.53	.61	.54	.79	.53	.65	.83	.20	.17
Openness to Experience	3.31(.64)	3.45(.61)	.57	.60	.43	.77	.57	.73	.80	.15	.17
Mean (of absolute values)			.56	.59	.55	.76	.50	.66	.78	.18	.23

Notes: All correlations in the table are significant at $p < .01$; all mean alphas and correlations are obtained after fisher r -to- z (back-)transformation; $N = 525$ for Study 1 (Community Sample) and $N = 309$ for Study 2 (Student Sample) except for temporal consistency ($N = 263$) and rater consistency ($N = 307$). T1 and T2 are respectively Time 1 and Time 2 measurement of the BHI in the Student Sample. When no information is provided on time of measurement in Student Sample, the data refer to T1.

agree) rated domain scales. The alpha reliabilities were .88 for Expressiveness, .87 for Preciseness, .84 for Verbal Aggressiveness, .83 for Questioningness, .84 for Emotionality, and .79 for Impression Manipulativeness.

From the student data obtained in Study 2, I selected other-rated HEXACO-PI-R (obtained at T1) and the following self-rated instruments: the Behavioral Activation Scale and the Behavioral Inhibition Scale (BAS/BIS; Gray, 1990; both obtained at T1 – wave 1), the Sensation Seeking Scale (SSS; Van den Berg & Feij, 2002; Zuckerman, 1979; T1 – wave 1), the Self-Concealment Scale (SCS; Larson & Chastain, 1990; T1 – wave 3), and the Experiences in Close Relations questionnaire (ECR; Brennan, Clark, & Shaver, 1998; T1 – wave 3). The HEXACO and SCS were rated on a 5-point (strongly disagree–strongly agree) scale; BAS and BIS were rated on a 4-point (strongly disagree–strongly agree) scale, and both the SSS and ECR Anxiety and Avoidance were rated on 7-point (strongly disagree–strongly agree) scales. The following alpha reliabilities were obtained for the other-rated HEXACO variables: .91 (H), .90 (E), .91 (X), .90 (A), .91 (C), and .89 (O) and the following alpha reliabilities were obtained for the self-ratings: .74 for BAS, .80 for BIS, .91 for the SSS, .82 for the SCS, .89 for ECR Anxiety, and .92 for ECR Avoidance.

3. Results

3.1. Main psychometric properties of the BHI

In Table 1, an overview of some of the main psychometric properties of the BHI in the two samples is provided. The alpha reliabilities ranged between .43 and .72; in both samples and Extraversion had the highest alpha reliability. In the student sample, temporal consistency (test–retest stability) and rater consistency (self–other agreement) were also investigated. Two months test–retest stability ranged between .71 (for Extraversion) and .79 (for Conscientiousness) with a mean of .76. Self–other agreement ranged between .39 (for Agreeableness) and .58 (for Emotionality) with a mean of .50. The convergent correlations of the BHI with the original HEXACO-PI-R ranged between .59 and .73 in the community sample and between .72 and .83 in the student sample. The maximum absolute discriminant correlations – that is, the maximum off-diagonal absolute correlations when the same order of BHI (rows) and HEXACO-PI-R (columns) variables is used – ranged between .15 and .21 and between .17 and .32 in the community and student samples respectively. The convergent correlations of the other-rated BHI with the other-rated HEXACO-PI-R in the student sample (not reported in Table 1) ranged between .71 (for other-rated Agreeableness) to .84 (for other-rated Conscientiousness) and the maximum

absolute discriminant correlations ranged between .09 and .39. Note that the larger values in the student sample are most likely due to the fact that the data was obtained concurrently, whereas in the community sample the HEXACO-PI-R data was obtained 3.5 years prior to the BHI data.

3.2. Comparison correlations BHI and HEXACO-PI-R

In Table 2, correlations between the HEXACO-PI-R (left-most columns) and the BHI (right-most columns) on the one hand with 24 different instruments (12 instruments in the community sample and 12 other instruments in the student sample) on the other hand are provided. Although from a content perspective, all of the correlations can be deemed interesting in their own right (e.g., see De Vries & Van Kampen, 2010; De Vries et al., 2013, for more information on the correlations in the community sample), here I focus on the difference in the size of the HEXACO-PI-R correlations on the one hand and the BHI correlations on the other. As can easily be seen in the table, almost all boldfaced ($\geq .30$) correlations involving the HEXACO-PI-R were higher than those involving the BHI, except for the correlation between Conscientiousness and CSI Preciseness in the community sample and the correlation between Honesty–Humility and the Sensation Seeking Scale in the student sample. The profile correlations – after Fisher r -to- z (back-)transformation of the correlations – between the BHI correlations on the one hand and the HEXACO-PI-R correlations were .96 in the community sample and .98 in the student sample, showing that the pattern of correlations was highly similar.⁶

To graphically depict the changes in validity when using the BHI instead of the HEXACO-PI-R, I inspected the ‘loss in validity’ as a function of effect size in the community sample and the student sample. The graph was obtained as follows: (1) I collected all absolute correlations $\geq .08$ and $\leq .82$ involving the HEXACO-PI-R; (2) at every .01 step, I collected all correlations involving the HEXACO-PI-R within a .20 range, e.g., between .08 and .28, between .09 and .29, between .10 and .30, ..., until .62 and .82; (3) within each .20 range, I calculated the mean fisher r -to- z transformed differences (e.g., validity loss) between the HEXACO-PI-R and BHI correlations; (4) to prevent individual data points to have a large effect, I obtained a ‘moving average’ of these mean differences, by taking the mean of five consecutive points. To give an example, I computed the mean of the mean differences of .08–.28, .09–.29, .10–.30, .11–.31, and .12–.32 and transformed this mean back to corre-

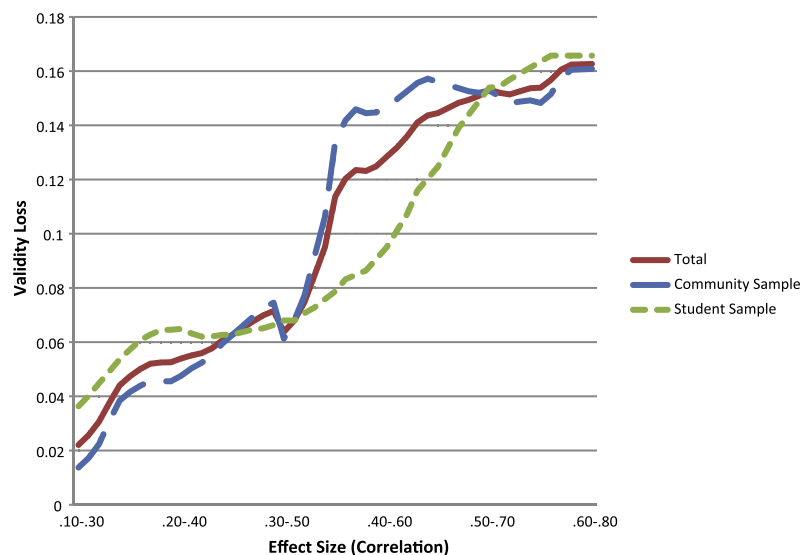
⁶ I also conducted a one-sided profile correlation analyses, in which all negative correlations were transformed into positive correlations except correlations with unequal sign (e.g., negative BHI correlations but positive HEXACO-PI-R correlations), which were transformed into negative correlations. The profile correlations were just slightly lower, i.e., .93 in the community sample, and .96 in the student sample.

Table 2

Comparison of correlations of HEXACO-PI-R and BHI with other constructs.

	HEXACO-PI-R (192 items)						BHI (24 items)					
	H	E	X	A	C	O	H	E	X	A	C	O
<i>Study 1: Community sample (N = 525)</i>												
5DPT Extraversion	-.07	-.08	.74	.03	.05	.13	-.12	-.08	.61	-.11	.04	.11
5DPT Neuroticism	-.06	.61	-.52	-.19	-.11	-.10	.04	.45	-.31	-.04	-.17	-.09
5DPT Absorption	.01	.21	.10	.03	.07	.51	-.06	.20	.19	.03	.03	.45
5DPT Orderliness	.04	.04	.01	-.03	.62	-.12	.06	.00	.04	.01	.55	-.14
5DPT Insensitivity	-.54	-.07	-.06	-.52	-.15	.13	-.40	-.01	-.14	-.31	-.23	.18
SRP Psychopathy	-.59	-.32	.10	-.27	-.19	.17	-.53	-.21	-.12	-.25	-.21	.18
CSI Expressiveness	-.07	-.07	.67	-.06	.05	.19	-.12	-.06	.62	-.17	.05	.20
CSI Preciseness	.03	-.25	.07	.15	.35	.14	.04	-.24	.06	.09	.41	.11
CSI Verbal Aggressiveness	-.42	-.09	.00	-.56	-.22	.01	-.40	-.03	-.12	-.45	-.26	.07
CSI Questioningness	-.20	-.13	.32	-.12	.08	.54	-.17	-.06	.21	-.15	-.01	.49
CSI Emotionality	-.03	.70	-.32	-.18	-.15	-.11	.03	.59	-.13	-.08	-.21	-.09
CSI Impression Manipulativeness	-.54	.12	.00	-.18	-.11	.05	-.40	.06	-.04	-.12	-.23	.14
<i>Study 2: Student sample (N = 307)</i>												
Other-rated Honesty–Humility	.54	.14	-.12	.28	.11	-.10	.44	.14	-.05	.23	.08	-.14
Other-rated Emotionality	.11	.70	-.17	-.10	.16	-.07	.18	.63	-.09	-.03	.12	-.04
Other-rated eXtraversion	-.11	-.17	.62	.02	-.12	.07	-.08	-.16	.53	-.04	-.10	.05
Other-rated Agreeableness	.22	.04	.00	.53	-.06	-.05	.16	.01	-.01	.43	-.04	-.06
Other-rated Conscientiousness	.25	.16	-.07	-.02	.68	-.17	.30	.12	-.07	.02	.57	-.21
Other-rated Openness to Experience	-.02	-.13	.05	.10	-.07	.69	-.07	-.07	.07	.02	.02	.57
Behavioral Activation Scale (BAS)	-.36	-.05	.27	-.30	-.08	.14	-.25	-.09	.18	-.28	-.14	.18
Behavioral Inhibition Scale (BIS)	.09	.64	-.32	-.14	.19	-.06	.22	.53	-.19	-.03	.09	-.05
Sensation Seeking Scale	-.35	-.35	.33	-.10	-.36	.47	-.36	-.33	.28	-.10	-.22	.42
Self-Concealment Scale	-.29	-.04	-.31	-.25	-.19	.06	-.26	.04	-.25	-.11	-.15	.05
ECR Anxiety	-.22	.40	-.35	-.19	-.06	.01	-.12	.38	-.29	-.05	-.09	-.01
ECR Avoidance	-.12	-.19	-.13	.03	-.24	.04	-.15	-.15	-.13	.10	-.22	-.01

Notes: $|r| \geq .30$ are boldfaced; in Study 1/Study 2 when $|r| > .11/.13$, $p < .01$; ECR = Experiences in Close Relations.

**Fig. 1.** Validity loss as a function of effect size.

lations to come up with an estimate of the .10–.30 validity loss. The final validity loss plots of the community sample and the student sample, and a sample-size $(N - 1)$ weighted average plot of these two samples ('total') is depicted in Fig. 1. As can be seen in this figure, when the correlations are weak (between .10 and .30), validity loss is relatively low as well (approximately .03). However, with strong correlations validity loss becomes progressively more pronounced (e.g., approximately .16 when correlations range between .60 and .80). Using all correlations from both samples, the average validity loss due to the use of the 24-item BHI instead of the 200-item HEXACO-PI-R was 22.5% with a standard deviation of 4.1%.

3.3. Re-estimation HEXACO-PI-R correlations from BHI correlations

To check whether it is possible to re-estimate the original HEXACO-PI-R correlations presented in Table 2, I corrected each of the correlations involving the BHI for unreliability using the correction for attenuation formula, i.e., $r_{yx(\text{corrected})} = r_{yx(\text{BHI scale})} \sqrt{\frac{r_{xx(\text{HEXACO-PI-R scale})}}{r_{xx(\text{BHI scale})}}}$. Note that the formula only corrects for the difference in reliability between the HEXACO-PI-R scale ($r_{xx(\text{HEXACO-PI-R scale})}$) and the BHI scale ($r_{xx(\text{BHI scale})}$); it does not completely correct for unreliability. Using this correction on the 72 BHI correlations in the student sample, I re-estimated each of the 72 original HEXACO-PI-R corre-

Table 3

The effects of different correlation correction procedures on the accuracy of corrected correlation estimates in the student sample (Study 2).

	Uncorrected	Corrected			
		Alpha (a)	Temporal (b)	Rater (c)	Weighted a + b + c
<i>M</i> (<i>SD</i>) of difference with original <i>r</i> 's	−.043(.066) ^a	.009(.074) ^c	−.025(.060) ^b	.031(.082) ^d	.002(.064) ^c
<i>t</i> -Test (and <i>p</i> -value) difference from zero	−5.47(<.01)	1.00(.32)	−3.50(<.01)	3.17(<.01)	0.27(.79)

Notes: All analyses involved 72 pairs of correlations. Values were obtained after fisher *r*-to-*z* (back-)transformation of correlations. A negative sign denotes a lower value for the (un-)corrected correlations; a positive sign denotes a higher value for the (un-)corrected correlations. Means and standard deviations with different superscripts differ significantly from each other at $p < .01$.

lations in four different ways: (a) using the BHI alpha reliabilities (e.g., 'Alpha'), (b) the test–retest stability correlations (e.g., 'Temporal'), (c) the self-other agreement correlations (e.g., 'Rater'), and (d) a sample-size ($N - 1$) weighted combination of the three consistency estimates (e.g., 'Weighted a + b + c').

Table 3 provides a summary on how much the uncorrected and corrected correlations differed from the original estimates. Negative means in Table 3 denote (un-)corrected correlations that were on average lower than the original HEXACO-PI-R correlations, and positive means denote (un-)corrected correlations that were on average higher than the original correlations. Two tests were performed: (a) I compared the uncorrected and corrected correlations among each other using a repeated measures ANOVA, and (b) I tested whether the estimates differed significantly from zero. The repeated measures ANOVA with Greenhouse-Geisser correction showed a significant overall difference between the (un-)corrected correlations ($F(1.35, 95.57) = 45.13, p < .001$). Additional post hoc tests using Bonferroni correction showed that all of the (un-)corrected correlations, except the alpha reliability corrected correlation and the weighted corrected correlation, differed significantly from each other. Subsequently, the *t*-test with zero test-value showed that all except these same two correction procedures differed significantly from zero. However, note that the alpha correction procedure resulted in a higher standard deviation than the weighted correction procedure. Consequently, the weighted combination of the three reliability estimates to correct the BHI correlations appears to result in the most accurate estimate of the original HEXACO-PI-R correlations.

4. Discussion and conclusions

Short personality scales are becoming increasingly popular, due to their usefulness in large-scale studies which leave scarce room for traditional lengthy personality instruments. Although there is widespread consensus that short scales should not be used for diagnostic purposes, the loss of alpha reliability does not seem to have very strong repercussions for the validity of short scales, especially when using a few well-chosen items per personality scale. There are currently a number of short Big Five instruments, but up until now, no short scale was available for the HEXACO model.⁷ In this research, I report the development of the Brief HEXACO Inventory (BHI), which represents the six HEXACO dimensions using 4 items per domain scale.

4.1. Use and misuse of consistency estimates

The findings show that although the BHI has relatively low alpha reliabilities, its domain scales show (1) adequate levels of test–retest stability, (2) adequate levels of self-other agreement, (3) high levels of convergent correlations with the HEXACO-PI-R

even though, in the case of the community sample, the measures were taken 3.5 years apart, (4) low levels of validity loss, especially in the .20–.40 range, and (5) high levels of convergence between the HEXACO-PI-R and BHI correlations with external variables after correction for attenuation using a weighted average of alpha, temporal, and rater consistency estimates. The findings underscore the misgivings several authors have about the importance of alpha reliability (e.g., [Sijtsma, 2009](#)) and underscore [McCrae et al.'s \(2011, p. 43\)](#) assertion that by itself "internal consistency [alpha] has little to do with [...] validity." The fact that the BHI scales have low alpha reliabilities does not seem to have major validity repercussions. One might even argue that higher alpha reliability levels found in other studies that have employed short personality scales (e.g., [Donnellan et al., 2006](#)) may be partly due to the use of 'narrower' trait items. This may, in turn, reduce the coverage of the personality domain and increase the chance of transient errors.

In this research, correcting the BHI correlations using self-other agreement (rater consistency) overestimated the original validity correlations on average, whereas correcting them for test–retest stability coefficients (temporal consistency) underestimated the validity correlations somewhat. Note, however, that the temporal consistency correction had the lowest standard deviation of all correction procedures. Although there was no significant difference between the alpha corrected estimates and the weighted corrected estimates, the most optimal correction procedure of the two in terms of the standard deviation and difference from zero was the weighted combination of alpha, temporal, and rater consistency estimates.⁸ Needless to say, future research needs to find out which of the above correction procedures works best in what kinds of circumstances. For instance, different outcomes of (combinations of) correction procedures are likely to be observed when choosing narrow over broad items (high versus low internal consistency), when using short instead of long test–retest periods (high versus low temporal consistency), and when using well-acquainted versus not well-acquainted other-ratings (high versus low rater consistency).

4.2. Adequacy of BHI

Readers may question whether the BHI is able to represent all of the HEXACO domains adequately. Especially with respect to Honesty–Humility, some may argue that it is harder to construct a valid measure of this construct, especially with few items, than it is to construct a valid measure of the other five constructs. That is, through impression management, people who are in reality low on Honesty–Humility may actually score high on Honesty–Humility. Although in certain circumstances (e.g., criminal investigations), people low on Honesty–Humility may be more likely to fake high levels of Honesty–Humility, until now, there is no

⁷ But please note that other researchers are currently working on another short questionnaire based on the HEXACO model. Please contact the author for more details.

⁸ However, care should be taken when using attenuation correction procedures. Correction for attenuation is known to increase sampling error, and may thus lead to biased parameters. Consequently, it should only be conducted in large-scale random samples, in which parameter estimates are highly accurate and sampling errors are minimal.

evidence that in conditions of anonymous responding people are more likely to fake Honesty–Humility than they are to fake on the other five personality scales (e.g., De Vries, Zettler, & Hilbig, *in press*; Lee et al., 2003). Furthermore, in samples of well-acquainted others, self-other agreement on Honesty–Humility is generally on par with self-other agreement on the other five personality scales (e.g., De Vries, Lee, & Ashton, 2008). In this study, self-other agreement, test–retest stability, and validity of BHI Honesty–Humility were highly comparable to the other BHI scales. In fact, self-other agreement on Honesty–Humility was higher than on Extraversion and Agreeableness, whereas the alpha reliabilities of Extraversion were on average higher, and alpha reliabilities of Agreeableness were on average lower than those of Honesty–Humility. Therefore, both in research settings (the community sample) and in developmental assessments (the student sample), BHI Honesty–Humility does not seem to function particularly worse than the other five constructs in terms of reliability and validity.

4.3. Further shortening?

Is further shortening of the HEXACO possible? I agree with Credé et al. (2012) that two items should probably be the minimum scale length when one wants to avoid resorting to complex formulated single items. But, unless complex statements are used, even two items are not enough to cover the breadth of each of the personality domains and moderate gains in validity have still been found when increasing the number of items (Credé et al., 2012). With four items per domain, it is possible to cover both the breadth of the personality domains and to include the high and low poles of these domains more adequately than with two items. With four such items, alpha reliabilities are still relatively low, but convergent validity and construct validity seem not to be affected to a great extent, even when a long period of time has elapsed, as was the case in the community sample. Consequently, especially for large-scale longitudinal studies, a one-time measurement of the 24-item BHI may not be too costly in terms of response time and may provide adequate validities even over periods spanning several years. For ESM studies, 24 items are probably too much to measure more than once, but then again, traits – but not states – need to be measured only once in ESM studies. The BHI shows relatively high levels of test–retest stability and consequently, when measured only once, the BHI may also be considered useful in ESM studies.

To summarize, the Brief HEXACO Inventory (BHI) is the first short questionnaire operationalization of the six-dimensional HEXACO personality model. With 24-items and four items per personality domain, the BHI combines both shortness and content breadth. Although the alpha reliabilities of the BHI scales are – as expected – low, the scales show relatively high levels of self-other agreement and test–retest validity, high levels of convergent validity and low levels of discriminant validity, and acceptable levels of validity loss. Although the BHI is not suitable for diagnostic assessment purposes and for studies investigating the incremental validity of personality, it is highly suitable for exploratory research in large-scale representative samples. In such samples, using the correction for attenuation formula based on a weighted average of alpha, temporal, and rater consistency estimates, researchers may even obtain a reasonable estimate of the effect size of the original HEXACO-PI-R. However, researchers are advised to follow-up exploratory research testing hypotheses using lengthier questionnaire operationalizations (e.g., 60- or 100-item versions; see www.hexaco.org) of the HEXACO model. Nonetheless, because the addition of Honesty–Humility offers added value in the prediction of socio-political attitudes and counterproductive behaviors

that are highly relevant in our current society, the BHI may be a useful tool in large-scale exploratory investigations.

Acknowledgements

Grateful acknowledgement is provided to Michael C. Ashton and Kibeom Lee for comments on an earlier version of the manuscript and to Joshua Tybur and Lisette Thooft for their help (back-)translating the BHI.

Appendix A. The Brief HEXACO Inventory (BHI)⁹

Instructions: Please indicate to what extent you agree with the following statements, using the following answering categories: 1 = strongly disagree, 2 = disagree, 3 = neutral (neither agree, nor disagree), 4 = agree, and 5 = strongly agree.

#	Dutch version	English version
1.	Ik kan lang naar een schilderij kijken.	I can look at a painting for a long time.
2.	Ik zorg dat dingen altijd op de juiste plek liggen.	I make sure that things are in the right spot.
3.	Ik blijf onaardig tegen iemand die gemeen was.	I remain unfriendly to someone who was mean to me.
4.	Niemand wil graag met mij praten.	Nobody likes talking with me.
5.	Ik ben bang om pijn te lijden.	I am afraid of feeling pain.
6.	Ik vind het moeilijk om te liegen.	I find it difficult to lie.
7.	Ik vind wetenschap saai.	I think science is boring.
8.	Ik stel ingewikkelde taken zo lang mogelijk uit.	I postpone complicated tasks as long as possible.
9.	Ik geef vaak kritiek.	I often express criticism.
10.	Ik leg gemakkelijk contact met vreemden.	I easily approach strangers.
11.	Ik maak me minder zorgen dan anderen.	I worry less than others.
12.	Ik ben benieuwd hoe je op een oneerlijke manier veel geld kan verdienen.	I would like to know how to make lots of money in a dishonest manner.
13.	Ik heb veel fantasie.	I have a lot of imagination.
14.	Ik werk erg nauwkeurig.	I work very precisely.
15.	Ik ben het snel met anderen eens.	I tend to quickly agree with others.
16.	Ik praat graag met anderen.	I like to talk with others.
17.	Ik kan prima in m'n eentje moeilijkheden overwinnen.	I can easily overcome difficulties on my own.
18.	Ik wil graag beroemd zijn.	I want to be famous.
19.	Ik houd van mensen met rare ideeën.	I like people with strange ideas.
20.	Ik doe vaak dingen zonder echt na te denken.	I often do things without really thinking.
21.	Zelfs als ik slecht behandeld word, blijf ik kalm.	Even when I'm treated badly, I remain calm.
22.	Ik ben zelden opgewekt.	I am seldom cheerful.
23.	Ik moet huilen bij trieste of	I have to cry during sad or

⁹ All reported data are based on the Dutch version. For the other-rated version of the BHI, please contact the author.

Appendix A (continued)

#	Dutch version	English version
	romantische films.	romantic movies.
24.	Ik heb recht op een speciale behandeling.	I am entitled to special treatment.

Scoring table BHI (recode scores of items followed with an 'R' as follows: 5→1, 4→2, 3→3, 2→4, 1→5): *Honesty–Humility*: 6 (Sincerity), 12R (Fairness), 18R (Greed Avoidance), 24R (Modesty); *Emotionality*: 5 (Fearfulness), 11R (Anxiety), 17R (Dependence), 23 (Sentimentality); *eXtraversion*: 4R (Social Self-esteem), 10 (Social Boldness), 16 (Sociability), 22R (Liveliness); *Agreeableness*: 3R (Forgiveness), 9R (Gentleness), 15 (Flexibility), 21 (Patience); *Conscientiousness*: 2 (Organization), 8R (Diligence), 14 (Perfectionism), 20R (Prudence); *Openness to Experience*: 1 (Aesthetic Appreciation), 7R (Inquisitiveness), 13 (Creativity), 19 (Unconventionality).

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jrp.2013.09.003>.

References

- Aronson, Z. H., Reilly, R. R., & Lynn, G. S. (2006). The impact of leader personality on new product development teamwork and performance: The moderating role of uncertainty. *Journal of Engineering and Technology Management*, 23(3), 221–247.
- 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.
- 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.
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., et al. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholinguistic studies in seven languages. *Journal of Personality and Social Psychology*, 86(2), 356–366.
- Ashton, M. C., Lee, K., Pozzebon, J. A., Visser, B. A., & Worth, N. C. (2010). Status-driven risk taking and the major dimensions of personality. *Journal of Research in Personality*, 44(6), 734–737.
- Becker, G. (2000). How important is transient error in estimating reliability? Going beyond simulation studies. *Psychological Methods*, 5(3), 370–379.
- Bernard, L. C., Walsh, R. P., & Mills, M. (2005). Ask once, may tell: Comparative validity of single and multiple item measurement of the Big-Five personality factors. *Counseling and Clinical Psychology Journal*, 2(1), 40–57.
- Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult attachment: An integrative overview. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 46–76). New York: Guilford Press.
- Burisch, M. (1997). Test length and validity revisited. *European Journal of Personality*, 11(4), 303–315.
- Cattell, R. B. (1973). *Personality and mood by questionnaire*. San Francisco: Jossey-Bass.
- Chmielewski, M., & Watson, D. (2009). What is being assessed and why it matters: The impact of transient error on trait research. *Journal of Personality and Social Psychology*, 97(1), 186–202.
- Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology*, 102(4), 874–888.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- 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 en Organisatie*, 22(3), 232–274.
- De Vries, R. E., Bakker-Pieper, A., Konings, F. E., & Schouten, B. (2013). The Communication Styles Inventory (CSI): A six-dimensional behavioral model of communication styles and its relation with personality. *Communication Research*, 40(4), 506–532.
- De Vries, R. E., & Born, M. Ph. (2013). De vereenvoudigde HEXACO persoonlijkheidsvragenlijst en een additioneel interstitieel proactiviteitsfacet [The simplified HEXACO personality inventory and an additional interstitial proactivity facet]. *Gedrag en Organisatie*, 26(2), 222–243.
- De Vries, A., De Vries, R. E., & Born, M. Ph. (2011). Broad versus narrow traits: Conscientiousness and honesty–humility as predictors of academic criteria. *European Journal of Personality*, 25(5), 336–348.
- De Vries, R. E., De Vries, A., De Hoogh, A., & Feij, J. (2009). More than the Big Five: Egoism and the HEXACO model of personality. *European Journal of Personality*, 23(8), 635–654.
- De Vries, R. E., Zettler, I., & Hilbig, B. E. (in press). Rethinking trait conceptions of social desirability scales: Impression Management as an expression of Honesty–Humility. *Assessment*. <http://dx.doi.org/10.1177/1073191113504619>.
- De Vries, R. E., Lee, K., & Ashton, M. C. (2008). The Dutch HEXACO Personality Inventory: Psychometric properties, self-other agreement, and relations with psychopathy among low and high acquaintanceship dyads. *Journal of Personality Assessment*, 90(2), 142–151.
- De Vries, R. E., & Van Kampen, D. (2010). The HEXACO and 5DPT models of personality: A comparison and their relationships with psychopathy, egoism, pretentiousness, immorality, and machiavellianism. *Journal of Personality Disorders*, 24(2), 244–257 [doi: 10.1521/pedi.2010.24.2.244].
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP Scales: Tiny-yet-effective measures of the Big Five Factors of Personality. *Psychological Assessment*, 18(2), 192–203.
- Duriez, B., Van Hiel, A., & Kossowska, M. (2005). Authoritarianism and social dominance in Western and Eastern Europe: The importance of the sociopolitical context and of political interest and involvement. *Political Psychology*, 26(2), 299–320.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102(4), 652–670.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504–528.
- Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition and Emotion*, 4(3), 269–288.
- 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.
- Hofstee, W. K. (1994). Who should own the definition of personality? *European Journal of Personality*, 8(3), 149–162.
- John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 461–494). New York: Guilford Press.
- Langford, P. H. (2003). A one-minute measure of the Big Five? Evaluating and abridging Shafer's (1999a) Big Five markers. *Personality and Individual Differences*, 35(5), 1127–1140.
- Larson, D. G., & Chastain, R. L. (1990). Self-concealment: Conceptualization, measurement, and health implications. *Journal of Social and Clinical Psychology*, 9(4), 439–455.
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39(2), 329–358.
- Lee, K., & Ashton, M. C. (2006). Further assessment of the HEXACO Personality Inventory: Two new facet scales and an observer report form. *Psychological Assessment*, 18(2), 182–191.
- Lee, K., Ashton, M. C., Ogunfowora, B., Bourdage, J. S., & Shin, K.-H. (2010). The personality bases of socio-political attitudes: The role of honesty–humility and openness to experience. *Journal of Research in Personality*, 44(1), 115–119.
- Lee, K., Ashton, M. C., Pozzebon, J. A., Visser, B. A., Bourdage, J. S., & Ogunfowora, B. (2009). Similarity and assumed similarity in personality reports of well-acquainted persons. *Journal of Personality and Social Psychology*, 96(2), 460–472.
- Lee, K., Ashton, M. C., Wiltshire, J., Bourdage, J. S., Visser, B. A., & Gallucci, A. (2013). Sex, power, and money: Prediction from the dark triad and honesty–humility. *European Journal of Personality*, 27(2), 169–184.
- Lee, K., Gizzarone, M., & Ashton, M. C. (2003). Personality and the likelihood to sexually harass. *Sex Roles*, 49(1–2), 59–69.
- 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.
- Nunnally, J. C. (1981). *Psychometric theory* (2nd ed.). New Delhi: Tata McGraw-Hill.
- Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin*, 21(2), 100–108.
- Paunonen, S. V., & Jackson, D. N. (1985). The validity of formal and informal personality assessments. *Journal of Research in Personality*, 19(4), 331–342.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203–212.
- Rushton, J. P., Bons, T. A., Ando, J., Hur, Y.-M., Irwing, P., Vernon, P. A., et al. (2009). A general factor of personality from multitrait-multimethod data and cross-national twins. *Twin Research and Human Genetics*, 12(4), 356–365.
- Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual-differences constructs. *Psychological Methods*, 8(2), 206–224.
- Sijsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107–120.
- Thorndike, R. L. (1951). Reliability. In E. F. Linquist (Ed.), *Educational measurement* (pp. 560–620). Washington, DC: American Council on Education.
- Van den Berg, P. T., & Feij, J. A. (2002). *SpanningsBehoefteLijst – selectieversie: Handleiding [Sensation Seeking Scale – selection version: Manual]*. Leiden: Psychologische Instrumenten Tests Services.
- Van Gelder, J.-L., & De Vries, R. E. (2012). Traits and states: Integrating personality and affect into a model of criminal decision making. *Criminology*, 50(3), 637–671.

- Van Kampen, D. (2012). The 5-Dimensional Personality Test (5DPT): Relationships with two lexically based instruments and the validation of the absorption scale. *Journal of Personality Assessment*, 94(1), 92–101.
- Williams, K. M., Paulhus, D. L., & Hare, R. D. (2007). Capturing the four-factor structure of psychopathy in college students via self-report. *Journal of Personality Assessment*, 88(2), 205–219.
- Woods, S. A., & Hampson, S. E. (2005). Measuring the Big Five with single items using a bipolar response scale. *European Journal of Personality*, 19(5), 373–390.
- Zettler, L., & Hilbig, B. E. (2010). Attitudes of the selfless: Explaining political orientation with altruism. *Personality and Individual Differences*, 48(3), 338–342.
- Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level of arousal*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.