



UvA-DARE (Digital Academic Repository)

The Dispositional Insight Test: Structure and Validity Across Cultures

de Kock, François S.; Fetvadjiev, Velichko H.; D'Urso, E. Damiano; de Vries, Reinout E.

DOI

[10.1080/00223891.2025.2520277](https://doi.org/10.1080/00223891.2025.2520277)

Publication date

2026

Document Version

Final published version

Published in

Journal of Personality Assessment

License

CC BY-NC-ND

[Link to publication](#)

Citation for published version (APA):

de Kock, F. S., Fetvadjiev, V. H., D'Urso, E. D., & de Vries, R. E. (2026). The Dispositional Insight Test: Structure and Validity Across Cultures. *Journal of Personality Assessment*, 108(1), 79-89. <https://doi.org/10.1080/00223891.2025.2520277>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

The Dispositional Insight Test: Structure and Validity Across Cultures

François S. de Kock¹ , Velichko H. Fetvadjev^{2,3,*} , E. Damiano D'Urso⁴  and Reinout E. de Vries⁵ 

¹Section for Organizational Psychology, University of Cape Town, South Africa; ²Department of Social Psychology, University of Amsterdam, The Netherlands; ³Workwell Research Unit, North-West University, South Africa; ⁴Independent Researcher, Amsterdam, The Netherlands; ⁵Department of Experimental and Applied Psychology, Vrije Universiteit, The Netherlands

ABSTRACT

The Dispositional Insight Test (DIT) is designed to assess a person's knowledge representations of personality and behavior expressions in different situational contexts. Despite its usefulness for professionals who regularly assess people, the factor structure of the DIT and measurement properties in different national populations have yet to be tested. Here, we report on the dimensionality and cross-cultural equivalence of the full and short Dispositional Insight Test (DIT and DIT-S), a novel measure of an important but understudied concept. Participants were students and their acquaintances in three countries from distinct world regions ($n=817$ in the Netherlands, 434 in New Zealand, and 437 in South Africa), who completed the test online. Multiple-group confirmatory factor analysis for categorical outcomes supported a higher order dimensional structure comprising three factors: trait induction, trait extrapolation, and trait contextualization. Invariance testing showed this structure to be equivalent across the three cultural groups. The DIT exhibited expected correlations with cognitive ability. Future cross-cultural research on the DIT and its correlates is invited.

ARTICLE HISTORY

Received 9 December 2024
Accepted 3 June 2025

Understanding the behavior of people is important to humans as a highly social species (Fiske & Taylor, 2017; Kenny, 2020). Insight into others' personalities might be especially relevant for professionals (e.g., psychologists, physicians, social workers, police officers) who routinely assess the personality of clients or other contacts as part of their work, but also for job applicants, team members, potential friends, and relationship partners. Research evidence also suggests that informant reports of personality can be more reliable than self-reports and predict some external criteria better (Balsis et al., 2015), although the differences are subtle and do not generalize across domains (Lee et al., 2024). What allows social perceivers to accurately infer the personality traits of other individuals? The characteristics of the “good judge” (Funder, 2012) of personality have received extensive research attention, and the emerging consensus is that cognitive factors seem to be the primary drivers of personality judgment accuracy. In particular, a person's insight into dispositional information (e.g., understanding the relations between personality traits, behaviors, and situations) could explain why some individuals find it easier to “read” others' personalities. We define dispositional insight—in line with earlier work (Christiansen et al., 2005; De Kock et al., 2015; De Vries et al., 2021)—as structured knowledge representations that enable information processing about personality traits and their behavioral expressions in different

situations. Research findings indicate that dispositional insight predicts the accuracy of perceivers' judgments of personality better than all other individual differences (for a review, see De Kock et al., 2020).

As mounting evidence shows that dispositional insight is an important feature of accurate judges, various measures of dispositional insight have been developed for use in research and industry practice (e.g., training and screening of raters). The Dispositional Insight Test (DIT; De Vries et al., 2021) is an ability-based and untimed test of insight into personality, behavior, and situations, comprising 78 multiple-choice items with the correct response for each item based on known empirical relationships (e.g., between behavioral expressions and traits, or among traits themselves) and subsequent keying by subject-matter experts. Various measurement-related questions need to be addressed, however, to advance the assessment of dispositional insight. That is, the main aim of this study is to advance the assessment of dispositional insight by describing the theoretical conceptualization, dimensionality, and cross-cultural measurement invariance (MI) of the DIT, for both the complete (DIT) and short form (DIT-S). Furthermore, to examine the instrument's nomological network across cultures, we assess its relationship with cognitive ability. Our evaluation relies on data collected in Dutch and English from three countries from distinct world

CONTACT François S. de Kock  francois.dekock@uct.ac.za  School of Management Studies, University of Cape Town, Rondebosch, Cape Town 7701, South Africa. *Affiliation at the time of data collection: Victoria University of Wellington, New Zealand. Data collection in New Zealand was supported by start-up grant 209547 to Velichko Fetvadjev at Victoria University of Wellington.

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/00223891.2025.2520277>. Contact the study authors for access to the DIT and DIT-S answer keys.

© 2025 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.

regions (Inglehart & Welzel, 2005): the Netherlands, New Zealand, and South Africa.

Dispositional insight: Conceptualization and measurement

The understanding of personality, of the ways personality manifests in behavior, and of the ways situations affect behavioral expression are seen as characteristics of the “good judge” (Funder, 1999). Building on these notions, Christiansen et al. (2005) conceptualized dispositional intelligence as an individual-difference construct relevant to the processing of dispositional information. In this conceptualization, dispositional insight requires the use of three declarative knowledge structures (De Kock et al., 2015, pp. 202–204): (a) “the ability to know how traits manifest themselves in behavior” (*trait induction*); (b) “the ability to identify situations that are relevant to different traits” (*trait contextualization*); and (c) “the understanding of how traits and their behavioral manifestations naturally covary” (*trait extrapolation*), also seen as the accuracy of a person’s implicit personality theory (Schneider, 1973).

Based on this conceptualization, three progressively more refined measures of dispositional insight have been developed. The Interpersonal Judgment Inventory (IJI; Christiansen et al., 2005) includes 45 items and has served as the basis for newer measures. The IJI was revised by De Kock et al. (2015) with longer subscales to allow for reliable facet-level scores to be computed for the components. The Revised Interpersonal Judgment Inventory (R-IJI) has 64 items and includes content relevant to the work context, in contrast to the IJI, which had items written with the college student context in mind. Building on this work, a new measure of dispositional insight was proposed (De Vries et al., 2021). The main three reasons for constructing a new measure were (a) a measure should be as generic as possible rather than being focused on students or the work context; (b) the development of previous measures did not attend to the systematic mapping of items onto personality facets—a desirable feature that would promote more balanced assessment of the complete content domain of insight into personality, behaviors, and situation-relevance of traits; and (c) considering the lexical evidence that supports a maximum of six cross-culturally replicable dimensions of personality (Ashton et al., 2004), it would be useful to construct an instrument that not only covers the Big Five lexical space, but also honesty-humility as an additional factor.

Construction of the Dispositional Insight Test

Against this background, the DIT was developed to measure people’s dispositional insight into the HEXACO traits (Lee & Ashton, 2004). Its construction followed a rational-empirical approach combined with an assessment of the items by four experts. In the rational-empirical stage, for the trait induction facet, items (adjectives) were selected based on several considerations: (a) there should be an equal number of positively and negatively keyed items (two adjectives for each trait); (b) the adjective should not be too tautological of the

domain name (e.g., not *emotional* for the domain emotionality); (c) the adjective should be relatively easy to understand; (d) the adjective should cover—preferably—different facets of the domain; and (e) the adjective should have high loadings on the construct of interest and low cross-loadings on other constructs, based on lexical research (Ashton et al., 2004) or other empirical evidence (e.g., egoism for honesty-humility; De Vries et al., 2009). For the trait extrapolation and trait contextualization items, we relied heavily on the content of the facets of the HEXACO-PI-R, which has very adequate psychometric properties concerning within-domain facet convergence and between-domain facet discrimination (Thielmann et al., 2020). For the extrapolation items, when provided with a behavior from a facet, we selected as the correct answer a behavior from another facet from the same domain and as the incorrect answers behaviors from other facets from different domains. In four cases, the items were reversed and the respondents needed to indicate which behavior was least likely. In those cases, the correct answer was a behavior selected from a facet from a different domain and the incorrect answers were behaviors from the same domain. Similarly, based on the HEXACO-PI-R facets and items (De Vries et al., 2016) and the trait activation model (Tett & Burnett, 2003), for the contextualization items, we selected situations that are theoretically associated with the trait activation of one of the following six facets from the corresponding HEXACO domains: fairness (from Honesty), sentimentality (Emotionality), sociability (Extraversion), forgiveness (Agreeableness), organization (Conscientiousness), and creativity (Openness). Incorrect answers were represented by behaviors associated with HEXACO facets from other (nontargeted) domains. In the expert phase, four personality experts (two final-year PhD students and two assistant professors) evaluated whether they understood the questions and the answers and whether the correct answer adequately represented the targeted construct. Based on their feedback, a few final modifications were made.

The DIT was developed in Dutch, while considering the previous, English-language measures of dispositional insight (Christiansen et al., 2005; De Kock et al., 2015). The DIT was translated into English by a fluent bilingual expert in the subject matter. The names of the hypothetical persons used in some of the items were adapted to the South African context (and were deemed to be familiar also in the New Zealand context). The English translation was examined by another fluent English speaker and expert in the field, who is an author of one of the earlier measures, and modifications were discussed in the research team. The translation procedure thus followed the outline of the committee approach, in line with current guidelines in cross-cultural assessment (Van de Vijver & Leung, 2021).

Structure

Although the DIT has been used in prior research (De Vries et al., 2021), its structural properties, including its cross-cultural replicability and validity, have not yet been investigated. According to its theoretical basis (Christiansen et al., 2005; De Vries et al., 2021), dispositional insight comprises the three

components of trait induction, extrapolation, and contextualization. In line with these earlier conceptualizations of dispositional insight we posit two plausible dimensional structures: (a) a one-factor model, where most of the variance between the elements of induction, extrapolation, and contextualization is shared and these are largely empirically indistinguishable as separate factors; and (b) a hierarchical model, with a higher order latent general factor influencing scores on the three lower order factors. Hierarchical models with only three lower order factors are statistically equivalent to three-factor models, where factors are allowed to be correlated (Eid et al., 2017). Hence, here we only consider the hierarchical model, which is in line with stratum theories of intelligence (e.g., Cattell-Horn-Carroll theory; see McGrew, 2005) that propose a psychometric structure for cognitive abilities with a clear hierarchical nature, with broad factors at a higher stratum affecting narrow factors at lower strata.

Research on the dimensionality of dispositional insight in the R-IJI has supported a factorial and hierarchical structure. Using a mixed sample ($N=321$) of managers and psychology students in South Africa, De Kock et al. (2017) found the best fit for a three-factor structure (compared to a one-factor model). Moreover, hierarchical confirmatory factor analysis (CFA) supported its hierarchical structure, with dispositional insight as a higher order construct underlying the three lower order factors.

Further research on the structure of dispositional insight is needed for two reasons. First, the extended coverage of the DIT with a basis in the HEXACO model calls for an evaluation of the DIT's dimensionality. Because the concepts of trait induction, extrapolation, and contextualization are the same in the DIT as in previous measures, we expect them to emerge as distinct factors. Second, a notable limitation of earlier research (De Kock et al., 2017) is that it used item parcels rather than individual item scores as factor indicators. There is a need to examine the structure using individual item scores, which allow for a more fine-grained and accurate assessment of measurement properties (Little et al., 2002).

Cross-cultural equivalence and nomological network

Research on dispositional insight has so far mostly been conducted in relatively homogeneous populations, typically with college students from the same university, in a single country, such as the United States (Christiansen et al., 2005), South Africa (De Kock et al., 2015, 2017), or the Netherlands (De Vries et al., 2021). However, for measures of dispositional insight to be useful in diverse contexts, we need to know if they are equivalent across cultures. Measurement equivalence exists when the properties of a measure concerning the target latent characteristic are the same across populations, which would be evident when knowledge of an individual's population membership tells us nothing about the expected score over and above knowledge of the characteristic being measured (Millsap, 2011). It is generally accepted in the field of cross-cultural research methodology that invariance of the measurement structure is a prerequisite to meaningfully comparing mean scores between

different populations (e.g., Church et al., 2011; Leitgöb et al., 2023; Meredith, 1993; Van de Vijver & Leung, 2021), even though different views have also been expressed (e.g., McCrae, 2015). As personality could have different meanings across cultures (Cheung et al., 2011), people's understanding of personality expression might also vary across cultures. Therefore, evidence of invariance would go a long way toward substantiating the use of the DIT in different cultures.

The invariance of dispositional insight measures has been examined across some subpopulations within the same country, such as managers versus psychology students (De Kock et al., 2017). However, the cross-cultural equivalence of dispositional insight has received no research attention so far. This is a consequential omission as potential comparisons of correlation patterns or mean scores between cultural groups might be vulnerable to measurement bias (Van de Vijver & Leung, 2021). Research on context effects (e.g., of culture and language) on people's insight into personality necessitates an examination of measurement equivalence.

Evidence from related fields allows us to expect some generalizability of the structure of dispositional insight across countries. For instance, ability measures in the domains of social and emotional intelligence (e.g., Ekermans et al., 2011; Libbrecht et al., 2014; Wang et al., 2012) have demonstrated structural equivalence across cultural groups. Measures of personality traits also show broadly equivalent factor structures across cultures and languages (McCrae et al., 2005; Thielmann et al., 2020), although higher levels of equivalence, such as metric or scalar invariance, are rarer (Cheung et al., 2011; Church, 2016; Dong & Dumas, 2020).

Finally, moving beyond internal psychometric properties, it is informative to examine the nomological networks of the DIT. Among potential correlates of dispositional insight, cognitive ability appears particularly relevant as evidence has been accumulated for a moderate-to-strong association of the two constructs (Christiansen et al., 2005; De Kock et al., 2015; De Vries et al., 2021). We offer a first investigation of this association in a cross-cultural context.

This study

To test the cross-cultural applicability of the DIT as an assessment tool for dispositional insight in nonclinical populations, this study examines the DIT's factor structure and measurement equivalence across three countries from distinct cultural regions of the world (Inglehart & Welzel, 2005): the Netherlands, New Zealand, and South Africa. We use item-level scores (compared to item parcels used in previous research) as input for our analyses. Specifically, we compare a one-factor model and a hierarchical model between countries, and we sequentially investigate levels of invariance based on specific parameter constraints, using procedures appropriate for ordered categorical outcomes (Wu & Estabrook, 2016). In addition to the complete instrument, we also examine a shorter version, which could be useful in situations when the test is part of a larger battery of tests and questionnaires.

Our research uses novel methods that amplify our contributions. Ability tests such as the DIT (with responses to

items scored as 0 or 1) render categorical (binary) outcomes. However, traditional invariance testing assumes continuous observed variables, whereas binary data are traditionally analyzed in an item-response-theory framework, usually focusing on differential item functioning. By contrast, we employ multiple-group categorical confirmatory factor analysis (MG-CCFA), which is well suited to model directly the categorical-outcome nature of the data from ability measures (Wu & Estabrook, 2016). When testing for MI, MG-CCFA has been shown to perform as well as or better than alternative item-level testing approaches such as multiple-group item-response theory (D'Urso et al., 2022).

In summary, this study seeks to test the measurement model of the DIT in the context of diverse, international, nonclinical populations. This is the first study to compare the measurement of people's insight into personality dynamics across cultures. This study can therefore shed valuable light on the measurement of abilities in the domain of personality cognition in multicountry research.

Method

Transparency and open science

The data files and R scripts are available on the Open Science Framework (OSF) at <https://doi.org/10.17605/OSF.IO/QDH3W>. In addition, the OSF page contains a printable version (MS Word file) of the test. The study was not preregistered.

Sample and procedure

Data were collected in the Netherlands, New Zealand, and South Africa in 2016–2017 in the framework of a study on the role of dispositional insight for the accuracy of personality ratings. This larger study used a self–other agreement design, collecting self- and peer-ratings from pairs of participants who knew each other. Data were excluded in cases of incomplete responses or when participants did not agree for their data to be used; in the data collection in New Zealand, only the data of complete dyads were retained. The data have not been analyzed at intermediate stages of data collection. The resulting samples of more than 400 participants per country were considered appropriate for examining invariance in structural equation modeling (SEM) according to common standards (e.g., Wolf et al., 2013).

The participants were students at universities in the three countries, as well as their friends and relatives recruited as peer-raters. The samples included 817 participants in the Netherlands, 434 in New Zealand, and 437 in South Africa. Data on ethnic background were collected only in New Zealand and South Africa, both of which have historically had a bi- or multiethnic composition. As a proxy for socioeconomic status, we collected data on parental education level; these data were also only collected in New Zealand and South Africa. The sample characteristics are presented in Table 1.

The questionnaires were completed online. Students in the Netherlands participated, together with their peers, voluntarily as part of a personality course and wrote a personality report on each other in which the DIT and International

Table 1. Demographic characteristics of the samples.

	The Netherlands	New Zealand	South Africa
Gender			
Females	576	299	289
Males	233	135	147
Other/not reported	8	—	1
Age in years, <i>M</i> (<i>SD</i>)	29.94 (14.52)	24.71 (12.12)	22.38 (6.60)
Ethnic group			
European/White	—	365	138
Māori	—	40	—
Black	—	—	127
Colored	—	—	97
Other	—	70	17
Parental education level (mother/father)			
None	—	11/17	6/10
School	—	95/102	130/114
Professional	—	123/117	112/106
University	—	148/140	175/189

Note. *N* = 1,688. The Netherlands *n* = 817, New Zealand *n* = 434, South Africa *n* = 437. Multiple identifications with ethnic group were possible and are relatively common in New Zealand.

Cognitive Ability Resource (ICAR; Condon & Revelle, 2014) scores were interpreted. Students in New Zealand received study credits for their participation. The peer-raters in New Zealand had a chance to enter a draw for supermarket vouchers to an approximate value of US\$18. Students with complete dyadic records in South Africa received payment to an approximate value of US\$10.

Measures

The DIT allows assessing a person's insight into HEXACO-based personality traits and behavioral expressions in different situations. The DIT consists of 78 items (13 per HEXACO factor), measuring three different domains: trait induction (24 items, 4 per HEXACO factor), trait extrapolation (30 items, 5 per HEXACO factor), and trait contextualization (24 items, 4 per HEXACO factor). The question stem for the induction items is the following: "Each of the words below describes one of the personality traits above. Think about each word and select the button that corresponds to the trait that you think best describes the word. The words can describe someone who scores high or low on the trait. Click the correct answer," and an example item is "insecure." Extrapolation is measured in three subscales, all of which require matching a described person to a trait or behavior. For example, the question stem of the first extrapolation subscale reads, "The following items contain descriptions of people, followed by traits that could be associated with these people. Click on the option that fits with the best answer." Finally, contextualization is measured in two subscales, both of which require relating a trait to a situation. For example, the instructions of the first extrapolation subscale read, "In the next section, you need to think about how situations allow traits to manifest in behavior. Each of the following asks you to choose the situation that is most relevant to the given trait. A situation is relevant to a trait if the situation can be expected to "provoke" the trait in some form of a behavioral response. For example, a situation that makes "aggression" relevant will include an opportunity

for someone to behave aggressively toward another. An example item follows to make this idea clear.” Multiple-choice items are employed and subsequently transformed into dichotomized responses; that is, incorrect responses are keyed 0 and correct responses are keyed 1. The DIT-S contains six items per domain (18 items total), which were selected based on the item-rest correlations and the rule that each of the six HEXACO domains should be covered by one item in each of the three purported subfactors of the DIT (De Vries et al., 2021). The complete instrument is available in the [Supplemental Material](#).

We measured cognitive ability using the 24-item version of the ICAR (Condon & Revelle, 2014). This instrument contains four subscales with six items each: Verbal Reasoning, Letters and Numbers, Matrix Reasoning, and Three-Dimensional Rotation. Supporting the validity of the ICAR, an even shorter (16-item) version of the ICAR was found to be very strongly related ($r > .80$) to the Wechsler Adult Intelligence Scale (4th ed. [WAIS-IV]; Young & Keith, 2020).

The Cronbach’s alphas of the DIT and the ICAR are presented in the left panel of [Table 2](#). The DIT’s alpha was .90 in the combined sample, with a range of .87 to .89 across the three countries. Although some of ICAR’s subscales had low alphas, the total scale’s alpha, which was relevant for this study, was .81 in the combined sample, with a range of .73 to .81 across the three countries.

Data analysis

To account for the categorical (i.e., dichotomous) type of data, we estimated all models using diagonally weighted least squares (DWLS) estimation, but the full-weight matrix was used to compare the mean- and variance-adjusted test statistics (default in *lavaan*; Rosseel, 2012).

First, we evaluated the dimensionality of the DIT and DIT-S by fitting, for each of the groups, (a) a higher order factor model and (b) a one-factor model. The former distinguishes three lower order factors (i.e., induction, extrapolation, and contextualization) from the higher order factor capturing general dispositional insight. The latter posits a general factor

structure where the three specific (lower order) factors are indistinguishable. To identify both the lower and higher order factors of the hierarchical model, we fixed their means and variances to 0 and 1, respectively. The fit for these group-specific models was evaluated using the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker–Lewis index (TLI). To imply satisfactory model fit, we followed common guidelines; that is, $RMSEA < .06$ and $CFI/TLI > .90$. In addition, we report the standardized root mean squared residual (SRMR), where a guideline of .08 is often used, but values below .10 are seen as indicating acceptable fit (Schermelleh-Engel et al., 2003). We did not use the χ^2 test for model evaluation because various studies have indicated that its performance is sample-dependent (Moshagen & Erdfelder, 2016; Rutkowski & Svetina, 2014).

After assessing the DIT dimensionality separately for each group and in the combined sample, we assessed measurement and structural invariance across countries. Specifically, we first tested invariance for the lower order factors (i.e., MI) followed by invariance for the higher order factor (i.e., structural invariance). For the lower order factors, we followed the MI steps and identification constraints discussed in Wu and Estabrook (2016). These authors have shown that, for dichotomous data—following configural invariance—invariance of thresholds, loadings, and intercepts cannot be evaluated separately and must be tested at once. After assessing MI for the lower order factors, we evaluated structural invariance (i.e., invariance of the higher order factor) by first assessing higher order weak invariance (i.e., invariance of the higher order factor loadings) followed by higher order metric invariance (i.e., invariance of the lower order factors’ intercepts). The criteria to evaluate the MI steps were the following: (a) we concluded configural invariance if $RMSEA < .06$ and $CFI > .90$; (b) we assessed invariance in the following steps by assessing the change in fit between the more restricted model and the less restricted model. Following Chen’s (2007) recommendations for large samples, we accepted invariance if one of two criteria was met: $\Delta CFI > -.01$ or $\Delta RMSEA < .01$ (i.e., to indicate noninvariance, both criteria would have to fail). Although $\Delta CFI > -.01$ is

Table 2. Cronbach’s alphas and descriptive statistics of the Dispositional Insight Test (DIT), the Short form of the Dispositional Insight Test (DIT-S), and the International Cognitive Ability Resource (ICAR; Condon & Revelle, 2014).

	Cronbach’s alpha				Country ^a M (SD)			
	NL	NZ	SA	Total	NL	NZ	SA	Total
DIT	.87	.89	.89	.90	56.51 (9.03)	44.86 (10.46)	48.22 (10.82)	51.14 (11.17)
Ind	.78	.81	.81	.83	15.98 (4.13)	11.36 (4.53)	12.16 (4.69)	13.71 (4.88)
Ext	.71	.68	.71	.73	21.78 (3.67)	17.70 (3.62)	18.87 (4.17)	19.89 (4.20)
Con	.72	.77	.79	.77	18.76 (3.07)	15.80 (3.96)	17.18 (4.04)	17.54 (3.81)
DIT-S	.77	.72	.77	.78	14.49 (2.65)	10.54 (2.79)	12.78 (3.36)	12.97 (3.32)
Ind	.68	.53	.62	.70	4.61 (1.50)	2.61 (1.35)	3.28 (1.62)	3.71 (1.73)
Ext	.62	.32	.64	.51	4.52 (0.91)	3.28 (0.81)	4.69 (1.45)	4.23 (1.21)
Con	.61	.62	.63	.63	5.36 (1.03)	4.65 (1.43)	4.82 (1.38)	5.02 (1.29)
ICAR	.73	.80	.81	.81	12.19 (4.09)	7.86 (4.41)	8.06 (4.55)	9.91 (4.79)
VR	.37	.52	.59	.56	3.93 (1.23)	2.79 (1.48)	2.66 (1.58)	3.28 (1.52)
LN	.63	.67	.65	.69	3.09 (1.72)	1.70 (1.63)	1.99 (1.65)	2.42 (1.79)
MR	.49	.52	.58	.52	2.71 (1.47)	2.32 (1.53)	2.39 (1.61)	2.52 (1.54)
R3D	.70	.75	.70	.76	2.47 (1.83)	1.06 (1.51)	1.03 (1.41)	1.70 (1.79)

Note. NL=The Netherlands; NZ=New Zealand; SA=South Africa; Ind=Induction; Ext=Extrapolation; Con=Contextualization; VR=Verbal Reasoning; LN=Letters and Numbers; MR=Matrix Reasoning; and R3D=Three-Dimensional Rotation.

^aSample sizes differ from [Table 1](#) due to combination of DIT and ICAR data: Total $N=1,611$; The Netherlands, $n=743$; New Zealand, $n=434$; South Africa, $n=434$.

sometimes used as sufficient to reject more restrictive models (Cheung & Rensvold, 2002), recent studies have proposed alternative cutoff values for Δ CFI (e.g., Rutkowski & Svetina, 2014). Furthermore, recent research suggested that, given the lack of consensus concerning fit indexes' cutoffs under all conditions and models, multiple fit criteria should be considered and reported (Putnick & Bornstein, 2016), notably for ordered-categorical data (D'Urso et al., 2022). Given the still fluid criteria for invariance in categorical models, especially for dichotomous data, we follow a more lenient approach to model comparison for invariance testing. Thus, we consider the scale to be noninvariant with respect to levels beyond configural invariance, if the change in CFI is < -0.01 and the change in RMSEA (a misfit measure) $> .01$.

Finally, we examined the correlations between the DIT and the ICAR using observed scores. Although, to our knowledge, the MI of the ICAR in intercultural contexts has not been examined, MI of the ICAR has been established among different age and gender groups (Young et al., 2019). For the purposes of this study, we are interested in the overall correlations of the DIT with ICAR as a first approach to the DIT's nomological network, but not in examining potential cross-cultural differences in the size of these correlations. As such, we abstain from analyzing the MI of the ICAR, which would entail a separate investigation and should be addressed in future research.

Software and packages

The models were analyzed using R (R Core Team, 2023). We used the *lavaan* package (Rosseel, 2012) for estimating the models and obtaining fit measures and the *semTools* package (Jorgensen et al., 2016) for specifying the MI models.

Results

Dimensionality assessment

The results concerning dimensionality are displayed in the top panel of Table 3. The higher order factor model provided an overall good fit across the three groups, with fit indexes for the combined sample, $\chi^2(2, 922) = 6,207.5$, RMSEA =

.026, CFI = .923, TLI = .921, SRMR = .066.¹ Additionally, the higher order factor model showed better goodness-of-fit results compared to the one-factor model, with fit indexes for the combined sample, $\chi^2(2, 925) = 9864.0$, RMSEA = .031, CFI = .883, TLI = .880, SRMR = .077. These results supported the notion of a hierarchical structure for the DIT for all groups, where a structure with a general dispositional insight that reflects the core of three lower level traits (induction, extrapolation, and contextualization) is a more accurate representation of the data in the three countries than a single-factor model.

Measurement and structural invariance testing

Fit statistics for both the measurement and structural invariance testing are displayed in Table 4. Overall, measurement and structural invariance held across the countries. Configural invariance held across groups as indicated by RMSEA = .018 and CFI = .954. Also, for all the subsequent steps, the Δ CFI < -0.01 was not complemented by a Δ RMSEA $> .01$, which suggests that invariance of the measurement (i.e., thresholds, loadings, and intercepts)² and structural (i.e., higher order factor loadings and intercepts) parameters held across groups. This result indicates that the lower level factors (i.e., induction, contextualization, and extrapolation) can be validly compared across groups. Furthermore, structural invariance results showed comparability for the higher level factor capturing general dispositional insight as both metric and scalar invariance were supported.

In addition, we scrutinized the higher and lower order factor loadings and the omega indexes of reliability, displayed in Tables 5 and 6, respectively.³ The omega indexes ranged from .792 to .897 across the lower and higher order factors. Most lower order factor loadings were above .30; however, some item loadings were low (below .30), which we noted with an asterisk in Table 5.

Short DIT

The results concerning the DIT-S's dimensionality are displayed in the bottom panel of Table 3. Like in the complete version of the DIT, the higher order factor model provided satisfactory fit both in the individual samples and in the combined sample, with fit indexes in the combined sample of $\chi^2(149) = 1239.4$, RMSEA = .040, CFI = .980, TLI = .977,

Table 3. Fit measures for the one- and higher order factor model within each group.

	One factor				Higher order factor			
	RMSEA	CFI	TLI	SRMR	RMSEA	CFI	TLI	SRMR
DIT								
NL	.023	.890	.887	.099	.018	.936	.934	.085
NZ	.023	.922	.920	.095	.018	.954	.952	.088
SA	.028	.891	.888	.105	.020	.942	.940	.093
Combined	.031	.883	.880	.077	.026	.923	.921	.066
DIT-S								
NL	.046	.888	.873	.113	.023	.973	.968	.076
NZ	.042	.933	.924	.099	.030	.965	.960	.084
SA	.062	.832	.809	.116	.044	.917	.904	.094
Combined	.063	.944	.937	.097	.040	.980	.977	.070

Note. RMSEA=root mean square error of approximation; CFI=comparative fit index; TLI=Tucker-Lewis Index; SRMR=standardized root mean squared residual; DIT=Dispositional Insight Test, DIT-S=Short form of the Dispositional Insight Test; NL=The Netherlands; NZ=New Zealand; SA=South Africa; Combined=combined sample.

¹Even though the SRMR values of the individual countries were above the commonly used guideline of .08, they were within the range of acceptable fit, below .10 (Schermelleh-Engel et al., 2003). Furthermore, the SRMR index is known to be sensitive to sample size and to be upward-biased (overestimating misfit; Shi et al., 2018), as highlighted by the lower SRMR in the combined sample. In combination with the other indexes, the model fit was thus judged to be fair both in the individual samples and in the combined sample.

²Although the change in CFI surpassed commonly used criteria (i.e., < -0.01), we consider thresholds, loadings, and intercepts to be noninvariant given the smaller changes in RMSEA as well as the combination of different model parameters tested at the same time.

³Following Flora's (2020) recommendations, the omega coefficient is appropriate to evaluate lower and higher order factors' reliability.

Table 4. Measurement invariance testing results for the Dispositional Insight Test (DIT).

	χ^2	df	p value	RMSEA	CFI	TLI	SRMR	Model comparison	Δdf	$\Delta RMSEA$	ΔCFI	ΔTLI	$\Delta SRMR$
Measurement invariance													
1. Configural	10429.3	8766	.0001	.018	.954	.942	.088		—	—	—	—	—
2. Thresholds, loadings, intercepts	11113.3	8910	.0001	.021	.925	.924	.092	2 vs. 1	144	.003	-0.018	-0.018	.004
Structural invariance													
3. Metric	11171.3	8914	.0001	.021	.923	.922	.094	3 vs. 2	4	.000	-0.002	-0.002	.001
4. Scalar	11212.9	8915	.0001	.021	.922	.921	.094	4 vs. 3	1	.000	-0.001	-0.001	.000

Note. RMSEA=root mean square error of approximation; CFI=comparative fit index; TLI=Tucker-Lewis Index; SRMR=standardized root mean squared residual.

Table 5. Standardized lower order factor loadings of the Dispositional Insight Test (DIT).

Induction		Extrapolation		Contextualization	
Item label	λ	Item label	λ	Item label	λ
Ind_1	.936	Ext_1	.313	Con_1	.008*
Ind_2	.529	Ext_2	.497	Con_2	.561
Ind_3	.654	Ext_3	.318	Con_3	-.027*
Ind_4	.573	Ext_4	.278*	Con_4	.255*
Ind_5	.454	Ext_5	.543	Con_5	.633
Ind_6	.712	Ext_6	.335	Con_6	.699
Ind_7	.745	Ext_7	.628	Con_7	.443
Ind_8	.480	Ext_8	.539	Con_8	.515
Ind_9	.617	Ext_9	.555	Con_9	.646
Ind_10	.450	Ext_10	.250*	Con_10	.193*
Ind_11	.767	Ext_11	.369	Con_11	.727
Ind_12	.659	Ext_12	-.026*	Con_12	.713
Ind_13	.598	Ext_13	.450	Con_13	.346
Ind_14	.556	Ext_14	.559	Con_14	.805
Ind_15	.752	Ext_15	.335	Con_15	.730
Ind_16	.676	Ext_16	.432	Con_16	.973
Ind_17	.785	Ext_17	.449	Con_17	.627
Ind_18	.190*	Ext_18	.241*	Con_18	.820
Ind_19	.644	Ext_19	.560	Con_19	.909
Ind_20	.112*	Ext_20	.488	Con_20	.694
Ind_21	.407	Ext_21	-.021*	Con_21	.620
Ind_22	.408	Ext_22	.839	Con_22	.842
Ind_23	.479	Ext_23	.426	Con_23	.399
Ind_24	.860	Ext_24	.737	Con_24	.215*
		Ext_25	.537		
		Ext_26	.647		
		Ext_27	.578		
		Ext_28	.548		
		Ext_29	.802		
		Ext_30	.797		
M	.59		.47		.56
SD	.19		.21		.27
Minimum	.11		-0.03		-0.03
Maximum	.94		.84		.97

Note. Full item statements corresponding to each item label are available in the [Supplemental Material](#).

*Items with standardized loading < .30.

SRMR = .070, and outperformed the one-factor model, with fit indexes in the combined sample of $\chi^2(152) = 2359.9$, RMSEA = .063, CFI = .944, TLI = .937, SRMR = .097. Thus, a higher order factor structure was supported for the DIT-S.

Table 7 displays the measurement and structural invariance results. Both measurement and structural invariance held for the DIT-S. Hence, also for the short DIT version, both the general dispositional insight and the lower order factors were comparable across countries. The factor loadings and reliabilities are displayed in the bottom panel of Table 6 and in Table 8. Compared to the complete DIT, lower order and higher order factor reliabilities were somewhat lower for the DIT-S, which is not surprising considering that a smaller set of items was used. Nevertheless, the reliability of the higher order factor (.842) was acceptable and would be sufficient for

Table 6. Standardized higher-order factor loadings and omega values.

	Induction	Extrapolation	Contextualization	General factor
DIT				
Loading	.769	.982	.908	—
Omega	.865	.781	.792	.897
DIT-S				
Loading	.747	.869	.957	—
Omega	.723	.673	.658	.842

Note. DIT=Dispositional Insight Test; DIT-S=Short form of Dispositional Insight Test.

assessment and individual decisions such as personnel selection (Cicchetti, 1994; Nunnally & Bernstein, 1994). Finally, factor loadings were greater than .40, with the only exception of one item (“You bump into a colleague who has—behind your back—blamed you for something that was not true to your supervisor”; Con_13), which was still fair at .36.

Correlations with ICAR and descriptive statistics

The correlations between the DIT and the ICAR are presented in Table 9. The general cross-cultural correlation was .48 and varied from .23 in the Netherlands to .46 in New Zealand. Finally, the means and standard deviations on the DIT and ICAR across the three countries are presented for descriptive purposes in the right panel of Table 2. Analyses of variance indicated that the Dutch sample scored higher than the other two samples on the DIT, $F(2, 1608) = 214, p < .001, \eta^2 = .21$; the DIT-S, $F(2, 1608) = 257, p < .001, \eta^2 = .24$; and the ICAR, $F(2, 1608) = 193, p < .001, \eta^2 = .19$.

Discussion

This study was the first to examine the dimensionality of the DIT across countries, as evidence of the hypothesized higher order factor model is important for future use of the measure in research and professional settings in different countries. The results demonstrated that the higher order three-factor solution of the DIT and DIT-S was the best fitting model and provided support for MI across three countries of distinct cultural regions: the Netherlands, New Zealand, and South Africa. Furthermore, in a first exploration of the cross-cultural nomological network of the DIT, we found that the DIT is correlated with cognitive ability across the three countries.

The support for the hierarchical three-factor solution for the DIT and DIT-S is consistent with earlier findings

Table 7. Measurement invariance testing results for the short version of the Dispositional Insight Test (DIT-S).

	χ^2	<i>df</i>	<i>p</i> value	RMSEA	CFI	TLI	SRMR	Model Comparison	Δdf	Δ RMSEA	Δ CFI	Δ TLI	Δ SRMR
Measurement invariance													
1. Configural	597.4	396	.0001	.030	.959	.953	.083		—	—	—	—	—
2. Thresholds, loadings, intercepts	678.4	420	.0001	.033	.948	.943	.086	2 vs. 1	24	.003	-0.012	-0.010	.003
Structural invariance													
3. Metric	686.1	424	.0001	.033	.947	.943	.086	3 vs. 2	4	.000	-0.001	.000	.000
4. Scalar	700.9	425	.0001	.034	.944	.940	.088	4 vs. 3	1	.001	-0.003	-0.003	.002

Table 8. Standardized lower-order factor loadings of the short version of the Dispositional Insight Test (DIT-S).

Induction		Extrapolation		Contextualization	
Item label	Λ	Item label	λ	Item label	λ
Ind_01	.911	Ext_25	.601	Con_13	.361
Ind_06	.681	Ext_26	.716	Con_15	.711
Ind_12	.678	Ext_27	.638	Con_16	.969
Ind_13	.578	Ext_28	.627	Con_17	.614
Ind_14	.526	Ext_29	.850	Con_19	.893
Ind_24	.786	Ext_30	.857	Con_20	.703

Note. Full item statements corresponding to each item label are available in the Supplemental Material.

distinguishing between induction, extrapolation, and contextualization of dispositional insight (Christiansen et al., 2005; De Vries et al., 2021) and with research indicating a hierarchical structure (De Kock et al., 2017). The hierarchical model balances theoretical complexity and practical utility in that it captures the contribution of individual items to the lower level factors and of the lower level factors to the general factor, at the same time affording the assessment of dispositional insight at both levels. Our results bolster the assessment of dispositional insight by establishing its structure in item-level analysis using categorical CFA, and extend the content of the DIT to the honesty-humility domain, captured in the HEXACO model. It is furthermore interesting to compare our factor-analytic results to typical results in personality research. Personality structure is complex, and CFA as a rule does not work at the item level in personality research, necessitating more flexible approaches to modeling (e.g., Marsh et al., 2010). By contrast, our results suggest that dispositional insight, at least when modeled using categorical CFA, exhibits a relatively simple factor structure.

The correlations between the DIT and cognitive ability that we found are broadly in line with previous research (Christiansen et al., 2005; De Vries et al., 2021), although some higher correlations have also been observed (De Kock et al., 2015). These findings suggest that key correlates of dispositional insight can be expected to be similar across cultures. It will be interesting for future research to examine the processes through which these constructs are linked and to expand further the nomological network of dispositional insight. For example, although the correlations with broad personality domains appear to be weak (De Vries et al., 2021), it would be interesting to examine the associations with more specific traits, such as empathy or narcissism.

The higher level of dispositional insight found in the Dutch sample compared to the other two samples invites a discussion, which must be seen as preliminary given the lack

of previous country comparisons. A large part of the Dutch sample was recruited from a course on personality and individual differences, where students had learned about traits and dimensional models of personality, whereas the students in the other two countries were recruited from different courses. Furthermore, students in this course anticipated written feedback on their DIT score, which might have formed a stronger incentive for accuracy. The higher scores of the Dutch sample appear to be in line with the proposition that dispositional insight might be trainable (Moran et al., 2022), yet without reducing the comparability of scores across samples.

Cross-cultural implications

Our findings on invariance have several implications for cross-cultural research and assessment. On a conceptual level, our findings suggest that the construct of dispositional insight is likely to have a relatively seamless transfer across cultures. Although presumed universal models are applicable to the three countries of our study (e.g., Milojev et al., 2013; Taylor & De Bruin, 2005; Thielmann et al., 2020), research from a local perspective in these countries has also identified various other, locally salient concepts of personality (e.g., De Raad & Barelds, 2008; Edwards et al., 2007; Fetvadjev et al., 2015). The finding of structural invariance of dispositional insight thus indicates that dispositional insight is probably not strongly sensitive to cross-cultural variations in implicit personality concepts; research in more cultures is certainly needed. Furthermore, higher levels of invariance such as metric and scalar invariance are often difficult to achieve in cross-cultural comparisons, particularly in a CFA framework and when using item-level data (Church et al., 2011; Van de Vijver & Leung, 2021). This makes the invariance observed in our study noteworthy, although we acknowledge the weaker support from the Δ CFI and the need for future research on change in fit indexes in categorical models, especially based on dichotomous data.

A practical implication of a cross-culturally equivalent measure of dispositional insight is that it can inform research on the trait-consistency, also referred to as traitedness, versus context-specificity of behavior across cultures (Church, 2009). There is evidence that cultures differ in the extent to which people tend to see behavior as predictable from traits and situationally consistent, although these differences in perceptions do not necessarily map onto differences in the predictability or consistency of actual behavior (Church, 2016; Fetvadjev et al., 2018). Future cross-cultural research stands to gain from examining the links between dispositional

Table 9. Correlations between the Dispositional Insight Test (DIT) and the International Cognitive Ability Resource (ICAR; Condon & Revelle, 2014).

	DIT	DIT: Ind	DIT: Ext	DIT: Con	ICAR	ICAR: VR	ICAR: LN	ICAR: MR	ICAR: R3D
DIT	—	.84/.86/.82	.86/.86/.86	.79/.86/.84	.23/.46/.42	.19/.45/.31	.16/.31/.36	.07/.31/.33	.17/.26/.21
DIT: Ind	.87	—	.54/.61/.53	.47/.57/.48	.20/.46/.45	.17/.41/.34	.15/.32/.34	.07/.29/.38	.14/.29/.25
DIT: Ext	.89	.64	—	.59/.67/.66	.21/.39/.30	.18/.38/.21	.15/.27/.26	.07/.28/.23	.15/.20/.17
DIT: Con	.84	.56	.68	—	.16/.34/.29	.13/.36/.22	.10/.21/.30	.04/.23/.20	.14/.18/.11
ICAR	.48	.47	.41	.35	—	.61/.72/.74	.71/.77/.77	.57/.68/.73	.71/.70/.67
ICAR: VR	.42	.41	.36	.32	.73	—	.33/.42/.46	.11/.33/.37	.29/.32/.31
ICAR: LN	.38	.36	.33	.28	.78	.47	—	.23/.34/.38	.26/.43/.34
ICAR: MR	.25	.25	.20	.17	.63	.28	.32	—	.19/.26/.35
ICAR: R3D	.34	.34	.30	.24	.74	.41	.41	.27	—

Note. Correlations below the diagonal are in the combined sample, and those above the diagonal are in the Netherlands, New Zealand, and South Africa, separated by slashes, respectively. Ind=Induction; Ext=Extrapolation; Con=Contextualization; VR=Verbal Reasoning; LN=Letters and Numbers; MR=Matrix Reasoning; R3D=Three-Dimensional Rotation. Correlations with $p < .001$ are presented in bold.

insight (both as an overall construct and its constituent components) and perceptions of traitedness.

From a methodological perspective, it is interesting to ask what aspects of the DIT are likely to have contributed to its measurement equivalence. First, the instrument begins with a detailed description of each of the HEXACO factors and contains detailed instructions in the various scales. Example questions with correct answers are provided for the scales of, arguably, the most complex task: trait contextualization. The provision of rich context is in line with item-writing guidelines aimed at the translatability of items (Brislin, 1986; International Test Commission, 2018). Second, several scales of the DIT are based on brief portraits of hypothetical persons; this is a format that has been shown to perform well in the cross-cultural assessment of abstract concepts such as values (Schwartz, 2003; Schwartz et al., 2001). Finally, translation procedures combining language and substantive expertise in a committee approach, like in this study, are generally seen as beneficial (International Test Commission, 2017; Van de Vijver & Leung, 2021; Walde & Völlm, 2023). Our findings suggest that combining these aspects in the construction and adaptation of instruments is likely to enhance their cross-cultural comparability.

Limitations

Despite several strengths, this study has limitations worth noting. First, although we collected data in three different countries, our convenience samples are not representative. We had mixed student and community samples—self-raters were students, and acquaintance raters were not all students—and these populations might differ in ability levels. Also, compared to national norms, most participants were well-educated as studies were conducted in universities. Further, more women than men participated in our study and the samples differed in their age composition. Additionally, as described earlier, the administration differed somewhat; the peer-report that was written based on the DIT and ICAR scores in the Dutch sample might have increased the students' motivation to do well on these two tests. Nevertheless, obtaining invariance despite sample differences in fact strengthens the case for an invariant structure. Future research stands to gain from examining the structure and predictive validity of the DIT in further cross-cultural data as well as participants from the general adult population (Simons et al., 2017).

Conclusion

Our findings have important implications for measuring dispositional insight in diverse populations of different countries. The full and short Dispositional Insight Test (DIT and DIT-S) showed favorable psychometric properties in Dutch, New Zealand, and South African samples. A hierarchical three-factor structure was observed in all country samples, and MI testing demonstrated metric and scalar invariance for both the complete and short forms. Furthermore, the DIT exhibited expected correlations with cognitive ability in all three countries. The DIT is a useful measure of a person's knowledge representations of personality and behavior expressions in different situational contexts. We invite future research examining the DIT and its correlates in more languages and cultures, with an eye to further expanding the cross-cultural applicability of this important concept.

Authors' contributions

François de Kock: Writing, data collection; Reinout de Vries: Writing, methodology, data collection; Velichko Fetvadjev: Writing, data collection; Damiano D'Urso: Formal analysis, data curation, writing. Data, headlines, and additional online materials are openly available at the project's Open Science Framework page.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

François S. de Kock  <http://orcid.org/0000-0003-3240-0570>
 Velichko H. Fetvadjev  <http://orcid.org/0000-0002-5475-5121>
 E. Damiano D'Urso  <http://orcid.org/0000-0002-2212-5440>
 Reinout E. de Vries  <http://orcid.org/0000-0002-4252-5839>

References

- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., Boies, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholinguistic studies in seven languages. *Journal of Personality and Social Psychology*, 86(2), 356–366. <https://doi.org/10.1037/0022-3514.86.2.356>
- Balsis, S., Cooper, L. D., & Oltmanns, T. F. (2015). Are informant reports of personality more internally consistent than self reports of personality? *Assessment*, 22(4), 399–404. <https://doi.org/10.1177/1073191114556100>

- Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137–164). Sage.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5
- Cheung, F. M., van de Vijver, F. J. R., & Leong, F. T. L. (2011). Toward a new approach to the study of personality in culture. *The American Psychologist*, 66(7), 593–603. <https://doi.org/10.1037/a0022389>
- Christiansen, N. D., Wolcott-Burnam, S., Janovics, J. E., Burns, G. N., & Quirk, S. W. (2005). The good judge revisited: Individual differences in the accuracy of personality judgments. *Human Performance*, 18(2), 123–149. https://doi.org/10.1207/s15327043hup1802_2
- Church, A. T. (2009). Prospects for an integrated trait and cultural psychology. *European Journal of Personality*, 23(3), 153–182. <https://doi.org/10.1002/per.700>
- Church, A. T. (2016). Personality traits across cultures. *Current Opinion in Psychology*, 8, 22–30. <https://doi.org/10.1016/j.copsyc.2015.09.014>
- Church, A. T., Alvarez, J. M., Mai, N. T. Q., French, B. F., Katigbak, M. S., & Ortiz, F. A. (2011). Are cross-cultural comparisons of personality profiles meaningful? Differential item and facet functioning in the Revised NEO Personality Inventory. *Journal of Personality and Social Psychology*, 101(5), 1068–1089. <https://doi.org/10.1037/a0025290>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, 43, 52–64. <https://doi.org/10.1016/j.intell.2014.01.004>
- D'Urso, E. D., De Roover, K., Vermunt, J. K., & Tijmstra, J. (2022). Scale length does matter: Recommendations for measurement invariance testing with categorical factor analysis and item response theory approaches. *Behavior Research Methods*, 54(5), 2114–2145. <https://doi.org/10.3758/s13428-021-01690-7>
- De Kock, F. S., Lievens, F., & Born, M. P. (2015). An in-depth look at dispositional reasoning and interviewer accuracy. *Human Performance*, 28(3), 199–221. <https://doi.org/10.1080/08959285.2015.1021046>
- De Kock, F. S., Lievens, F., & Born, M. P. (2017). A closer look at the measurement of dispositional reasoning: Dimensionality and invariance across assessor groups. *International Journal of Selection and Assessment*, 25(3), 240–252. <https://doi.org/10.1111/ijsa.12176>
- De Kock, F. S., Lievens, F., & Born, M. P. (2020). The profile of the 'Good Judge' in HRM: A systematic review and agenda for future research. *Human Resource Management Review*, 30(2), 100667. <https://doi.org/10.1016/j.hrmr.2018.09.003>
- De Raad, B., & Barelids, D. P. H. (2008). A new taxonomy of Dutch personality traits based on a comprehensive and unrestricted list of descriptors. *Journal of Personality and Social Psychology*, 94(2), 347–364. <https://doi.org/10.1037/0022-3514.94.2.347>
- De Vries, R. E., Barends, A. J., & De Kock, F. S. (2021). Dispositional insight: Its relations with HEXACO personality and cognitive ability. *Personality and Individual Differences*, 173, 110644. <https://doi.org/10.1016/j.paid.2021.110644>
- De Vries, R. E., De Vries, A., De Hoogh, A. H. B., & Feij, J. A. (2009). More than the Big Five: Egoism and the HEXACO model of personality. *European Journal of Personality*, 23(8), 635–654. <https://doi.org/10.1002/per.733>
- De Vries, R. E., Tybur, J. M., Pollet, T. V., & Van Vugt, M. (2016). Evolution, situational affordances, and the HEXACO model of personality. *Evolution and Human Behavior*, 37(5), 407–421. <https://doi.org/10.1016/j.evolhumbehav.2016.04.001>
- Dong, Y., & Dumas, D. (2020). Are personality measures valid for different populations? A systematic review of measurement invariance across cultures, gender, and age. *Personality and Individual Differences*, 160, 109956. <https://doi.org/10.1016/j.paid.2020.109956>
- Edwards, S., McCreanor, T., & Moewaka-Barnes, H. (2007). Maori family culture: A context of youth development in Counties/Manukau. *Kōtuitui: New Zealand Journal of Social Sciences Online*, 2(1), 1–15. <https://doi.org/10.1080/1177083X.2007.9522420>
- Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, 22(3), 541–562. <https://doi.org/10.1037/met0000083>
- Ekermans, G., Saklofske, D. H., Austin, E., & Stough, C. (2011). Measurement invariance and differential item functioning of the Bar-On EQ-i: S measure over Canadian, Scottish, South African and Australian samples. *Personality and Individual Differences*, 50(2), 286–290. <https://doi.org/10.1016/j.paid.2010.10.004>
- Fetvadjev, V. H., Meiring, D., Van de Vijver, F. J. R., Nel, J. A., & Hill, C. (2015). The South African Personality Inventory (SAPI): A culture-informed instrument for the country's main ethnocultural groups. *Psychological Assessment*, 27(3), 827–837. <https://doi.org/10.1037/pas0000078>
- Fetvadjev, V. H., Meiring, D., Van de Vijver, F. J. R., Nel, J. A., Sekaja, L., & Laher, S. (2018). Personality and behavior prediction and consistency across cultures: A multimethod study of Blacks and Whites in South Africa. *Journal of Personality and Social Psychology*, 114(3), 465–481. <https://doi.org/10.1037/pspp0000129>
- Fiske, S. T., & Taylor, S. E. (2017). *Social cognition: From brains to culture* (3rd ed.). Sage.
- Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, 3(4), 484–501. <https://doi.org/10.1177/2515245920951747>
- Funder, D. C. (1999). Chapter 6—Moderators of accuracy. In D. C. Funder (Ed.), *Personality judgment: A realistic approach to person perception* (pp. 137–168). Academic Press. <https://doi.org/10.1016/B978-012269930-6/50007-1>
- Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21(3), 177–182. <https://doi.org/10.1177/0963721412445309>
- Inglehart, R., & Welzel, C. (2005). *Modernization, cultural change, and democracy: The human development sequence*. Cambridge University Press.
- International Test Commission. (2017). *The ITC guidelines for translating and adapting tests* (2nd ed.). www.InTestCom.org
- International Test Commission. (2018). *ITC guidelines for the large-scale assessment of linguistically and culturally diverse populations*. www.InTestCom.org
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., Rosseel, Y., Miller, P., Quick, C., Garnier-Villareal, M., Selig, J., Boulton, A., Preacher, K., Coffman, D., Rhemtulla, M., Robitzsch, A., Enders, C., Arslan, R., Clinton, B., Panko, P., Merkle, E., Chesnut, S., Byrnes, J., ... Vanbrabant, L. (2016). Package 'semTools'.
- Kenny, D. A. (2020). *Interpersonal perception: The foundation of social relationships* (2nd ed.). Guilford.
- 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., De Vries, R. E., & Ashton, M. C. (2024). Self/observer agreement in personality assessment by observers' relationship types. *Journal of Research in Personality*, 113, 104529. <https://doi.org/10.1016/j.jrp.2024.104529>
- Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., Jak, S., Meitinger, K., Menold, N., Muthén, B., Rudnev, M., Schmidt, P., & van de Schoot, R. (2023). Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social Science Research*, 110, 102805. <https://doi.org/10.1016/j.ssresearch.2022.102805>
- Libbrecht, N., Beuckelaer, A. D., Lievens, F., & Rockstuhl, T. (2014). Measurement invariance of the Wong and Law Emotional Intelligence Scale scores: Does the measurement structure hold across Far

- Eastern and European countries? *Applied Psychology*, 63(2), 223–237. <https://doi.org/10.1111/j.1464-0597.2012.00513.x>
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the Big Five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22(3), 471–491. <https://doi.org/10.1037/a0019227>
- McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review*, 19(2), 97–112. <https://doi.org/10.1177/1088868314541857>
- McCrae, R. R., Terracciano, A., & 78 Members of the Personality Profiles of Cultures Project. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, 88(3), 547–561. <https://doi.org/10.1037/0022-3514.88.3.547>
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 136–181). Guilford.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/BF02294825>
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge/Taylor & Francis Group.
- Milojević, P., Osborne, D., Greaves, L. M., Barlow, F. K., & Sibley, C. G. (2013). The Mini-IPIP6: Tiny yet highly stable markers of Big Six personality. *Journal of Research in Personality*, 47(6), 936–944. <https://doi.org/10.1016/j.jrp.2013.09.004>
- Moran, J. B., Perry, L. M., Hoerger, M., & Murray, D. R. (2022). Dispositional intelligence of the Five-Factor Model as a learning outcome in an undergraduate personality course. *Scholarship of Teaching and Learning in Psychology*, 8(4), 404–408. <https://doi.org/10.1037/stl0000315>
- Moshagen, M., & Erdfelder, E. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 54–60. <https://doi.org/10.1080/10705511.2014.950896>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- R Core Team. (2023). *R* (4.3.1): A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>
- Rossee, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31–57. <https://doi.org/10.1177/0013164413498257>
- Schermelh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, 8(2), 23–74.
- Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin*, 79(5), 294–309. <https://doi.org/10.1037/h0034496>
- Schwartz, S. H. (2003). A proposal for measuring value orientations across nations. In *Questionnaire Development Package of the European Social Survey* (pp. 259–319). <https://www.europeansocialsurvey.org>
- Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., Harris, M., & Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology*, 32(5), 519–542. <https://doi.org/10.1177/0022022101032005001>
- Shi, D., Maydeu-Olivares, A., & DiStefano, C. (2018). The relationship between the standardized root mean square residual and model misspecification in factor analysis models. *Multivariate Behavioral Research*, 53(5), 676–694. <https://doi.org/10.1080/00273171.2018.1476221>
- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123–1128. <https://doi.org/10.1177/1745691617708630>
- Taylor, N., & De Bruin, G. (2005). *Basic Traits Inventory: Technical manual*. Jopie van Rooyen & Partners SA.
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *The Journal of Applied Psychology*, 88(3), 500–517. <https://doi.org/10.1037/0021-9010.88.3.500>
- Thielmann, I., Akrami, N., Babarović, T., Belloch, A., Bergh, R., Chirumbolo, A., Čolović, P., de Vries, R. E., Dostál, D., Egorova, M., Gnisci, A., Heydasch, T., Hilbig, B. E., Hsu, K.-Y., Izdebski, P., Leone, L., Marcus, B., Mededović, J., Nagy, J., ... Lee, K. (2020). The HEXACO–100 across 16 languages: A large-scale test of measurement invariance. *Journal of Personality Assessment*, 102(5), 714–726. <https://doi.org/10.1080/00223891.2019.1614011>
- Van de Vijver, F. J. R., & Leung, K. (2021). *Methods and data analysis for cross-cultural research*. (V. H. Fetvadjev, J. He, & J. R. J. Fontaine (Eds.), 2nd ed.). Cambridge University Press. <https://doi.org/10.1017/9781107415188>
- Walde, P., & Völlm, B. A. (2023). The TRAPD approach as a method for questionnaire translation. *Frontiers in Psychiatry*, 14, 1199989. <https://doi.org/10.3389/fpsy.2023.1199989>
- Wang, C., Kim, D.-H., & Ng, K.-M. (2012). Factorial and item-level invariance of an emotional intelligence scale across groups of international students. *Journal of Psychoeducational Assessment*, 30(2), 160–170. <https://doi.org/10.1177/0734282911412543>
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, 76(6), 913–934. <https://doi.org/10.1177/0013164413495237>
- Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, 81(4), 1014–1045. <https://doi.org/10.1007/s11336-016-9506-0>
- Young, S. R., & Keith, T. Z. (2020). An examination of the convergent validity of the ICAR16 and WAIS-IV. *Journal of Psychoeducational Assessment*, 38(8), 1052–1059. <https://doi.org/10.1177/0734282920943455>
- Young, S. R., Keith, T. Z., & Bond, M. A. (2019). Age and sex invariance of the International Cognitive Ability Resource (ICAR). *Intelligence*, 77, 101399. <https://doi.org/10.1016/j.intell.2019.101399>