Measuring Intellectual Curiosity across Cultures: Validity and Comparability of a New Scale in Six Languages

Matthias Bluemke, Lukas Engel, David J. Grüning & Clemens M. Lechner

Routledge
Taylor & Francis Group

Measuring Intellectual Curiosity across Cultures: Validity and Comparability of a New Scale in Six Languages

Matthias Bluemke[1] [iD], Lukas Engel[1,2] [iD], David J. Grüning[1,3] [iD] and Clemens M. Lechner[1] [iD]

[1]GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany; [2]University of Magdeburg, Germany; [3]University of Heidelberg, Germany

ABSTRACT

Intellectual curiosity—the tendency to seek out and engage in opportunities for effortful cognitive activity—is a crucial construct in educational research and beyond. Measures of intellectual curiosity vary widely in psychometric quality, and few measures have demonstrated validity and comparability of scores across multiple languages. We analyzed a novel, six-item intellectual curiosity scale (ICS) originally developed for cross-national comparisons in the context of the OECD's Programme for the International Assessment of Adult Competencies (PIAAC). Samples from six countries representing six national languages (U.S. Germany, France, Spain, Poland, and Japan; total $N=5,557$) confirmed that the ICS possesses very good psychometric properties. The scale is essentially unidimensional and showed excellent reliability estimates. On top of factorial validity, the scale demonstrated strict measurement invariance across demographic segments (gender, age groups, and educational strata) and at least partial scalar invariance across countries. As per its convergent and divergent associations with a broad range of constructs (e.g., Open-Mindedness and other Big Five traits, Perseverance, Sensation Seeking, Job Orientations, and Vocational Interests), it also showed convincing construct validity. Given its internal and external relationships, we recommend the ICS for assessing intellectual curiosity, especially in cross-cultural research applications, yet we also point out future research areas.

Curiosity, the tendency to seek out and pursue novel stimuli and challenging experiences in the environment (e.g., Kashdan et al., 2004), is an essential human disposition that resides at the intersection of intellectual ability and a general interest in learning experiences and stimulation. Maslow (1943) called it a central human motivation, and Peterson and Seligman (2004) deemed it a universal human strength that supports well-being. An important *facet* of the broader trait curiosity, on which we focus in this paper, is *intellectual curiosity* (IC), the tendency "to seek out, engage in, enjoy, and pursue opportunities for *effortful cognitive activity*" (von Stumm et al., 2011, p. 577).

IC has been termed the "third pillar" of academic performance (von Stumm et al., 2011; also see Orcutt & Dringus, 2017; Powell & Nettelbeck, 2014; von Stumm & Ackerman, 2013) showing relevant associations with cognitive development (Trudewind, 2000), academic (von Stumm et al., 2011) and non-formal education (Gorges et al., 2017), and professional growth (Mussel et al., 2012). IC has relevance for people's exploration of social contexts (Kashdan et al., 2020) and transcends disciplinary borders, for example, into pedagogy (e.g., impulse and attention control in children, Piotrowski et al., 2014; and students' scientific thinking, Zagumny, 2016, 2018) and computer science (e.g., perceived digital competence, Grüning & Lechner, 2023). Its diverse network of associations (see e.g., for creativity, Hardy et al., 2017; McCrae, 1987; or self-regulation

strategies, Lauriola et al., 2015; Wiggin et al., 2019) solidifies its position as one of the central and narrowly assessed personality traits of the human psyche.

IC derives from a rich research history, drawing on diverse assessment approaches and varying conceptualizations: initially perceived in behavioristic terms (Berlyne, 1954), it was later conceived as a (unitary) domain in personality inventories (Cacioppo et al., 1984; Goff & Ackerman, 1992). In current personality frameworks, it is mostly viewed as a facet woven into a comprehensive network of personality traits and underlying facets (Costa & McCrae, 1992; Johnson, 2014; Lee & Ashton, 2004; Soto & John, 2009). IC is one of three facets of the Open-Mindedness domain (also known as "Openness to Experience" or "Intellect") in the Big Five Inventory-2 (BFI-2; Soto & John, 2017).

The diverse research history has led to different prominent IC definitions. For instance, in the tradition of Kang et al. (2009) and Litman (2010), IC has been construed with a focus on its epistemic dimension as the "need or desire for knowledge, information, or the exploration of academic environments" (Grossnickle, 2016, p. 27). This phrasing comes without references to cognitive *effort* or *deprivation* sensitivity, both used in Kashdan et al. (2004) broader conceptualization. This—and further definitional variation laid out by Grossnickle (2016)—showcases that, despite high research efforts, to date, there is no agreement among scientists about how to conceptualize and best assess IC: The

research field is plagued with meandering between various curiosity objects that one might be willing to consider and distinguish on the one hand (e.g., physical, perceptual, social, epistemic, and intellectual curiosity; see Grossnickle, 2016) and the similarity of their motivational underpinning resulting in conceptual overlap on the other hand. As we will outline below, the approaches to measuring IC vary accordingly: Similar labels are used for rather disparate operationalizations ("jingle fallacy"), and similar concepts are investigated under different labels ("jangle fallacy").

All this has led to a highly problematic proliferation of measures. No specific measure of IC has superseded others or emerged as a gold standard, and this is especially true for cross-cultural applications. To make IC the human universal that it is supposed to be, researchers need to reduce conceptual disparity and rule out that cultural specificities hamper group comparison of scale scores. With the help of multinational data collected in six OECD countries, we answer the question of whether a new six-item scale allows for reliable, valid, and psychometrically comparable assessment of IC across the dominant languages in these countries, thereby pointing to the new scale as a valuable frontrunner in the race for a standard measure in large-scale assessment contexts.

Measurement of intellectual curiosity and related concepts

Before we describe the new scale that may serve as the—currently lacking—reference standard for measuring IC in international contexts, let us briefly review relevant traditions for measuring IC and related concepts, despite this list being far from exhaustive.

Big Five/Big Six personality trait-and-facet inventories

Prominent personality inventories typically house at least one facet underlying the domain of *Openness (to Experience)* that pertains to IC. Respective facets are either labeled as *Ideas* (NEO-PI-R; Costa & McCrae, 1992; BFI: Soto & John, 2009) with items such as "Is curious about many different things," or *Intellect* (IPIP-NEO-120; Johnson, 2014) with items such as "Like to solve complex problems" and "Avoid philosophical discussions," or *Inquisitiveness* (HEXACO-PI; Lee & Ashton, 2004; IPIP-HEXACO; Ashton & Lee, 2007) with exemplary items like "Enjoy intellectual games" and "Have excellent ideas." At times, the facets are specifically labeled as *Intellectual Curiosity* (BFI-2; Soto & John, 2017) and their items target at "Is complex, a deep thinker" and "Is curious about many different things," whereas in other inventories facets border marginally on IC such as *Excitement Seeking* (i.e., in NEO-PI-R's and IPIP-NEO-120's *Extraversion*) and items like "Seek adventure" (a general excitement to discover novel stimuli).

Epistemic Curiosity

Epistemic curiosity (EC) has been conceptualized as the desire for knowledge that motivates individuals to learn new ideas, eliminate information gaps, and solve intellectual problems (Litman, 2008). According to Grossnickle (2016),

IC is an overlapping, though more nuanced concept of EC which is why respective scales (e.g., ECS; Litman & Spielberger, 2003; and CFDS; Litman & Jimerson, 2004) include IC-like characteristics. For example, ECS items describe interests in solving problems (e.g., "Interested in trying to solve a riddle") and intellectual reasoning to gain deeper insight (e.g., "Enjoy discussing abstract concepts").

Need for Cognition

The *Need for Cognition Scale* (NfC, Cacioppo et al., 1984) captures the satisfaction elicited by mental processes of thinking. NfC-items that are reminiscent of the definition of IC (von Stumm et al., 2011) reflect a characteristic interest in problem-solving (e.g., "I prefer my life to be filled with puzzles that I must solve"), a preference for complexity (e.g., "I would prefer complex to simple problems"), and joy derived from abstract thinking and spending mental effort (e.g., "Thinking is not my idea of fun"). It has been argued that NfC may be measuring essentially the same intelligence-related personality characteristic as typical intellectual engagement (Woo et al., 2007).

Typical Intellectual Engagement

The *Typical Intellectual Engagement* inventory (TIE; Goff & Ackerman, 1992) targets the intellectual effort routinely spent on different tasks—as opposed to maximum intellectual performance (IQ). Three TIE scales assess reading, abstract thinking, and problem-directed thinking with items that address IC characteristics. For instance, the items reflect one's enthusiasm about solving relevant problems (e.g., "I prefer my life to be filled with puzzles I must solve") and being drawn to novelty (e.g., "I prefer activities I've never tried to ones I know I will enjoy").

Deprivation Sensitivity and Joyous Exploration

To disentangle multiple domains of curiosity Kashdan et al. (2020) recently presented a new, multifaceted instrument. They conceptualized curiosity in a multidimensional way and included, among others, two characteristics reminiscent of IC, namely *Deprivation Sensitivity* and *Joyous Exploration*. This distinction was first developed similarly by Litman (2008) who attempted to explain EC as the interplay of wanting- and liking-aspects of information-seeking. The dimensions were "deprivation type of curiosity" (EC-D; discomfort arising from perceived lack of information and unsolved puzzles, instigating uncertainty reduction even if effortful) and "interest" (EC-I; seeking intrinsic pleasure through learning). Kashdan et al. (2018, 2020) repackaged the items with little modification as part of the newer 5DC scale, while the concept definitions were slightly adjusted. The first factor is now defined epistemologically, namely as "being aware of information you do not know, want to know, and devote considerable effort to uncover" (p. 1). Two exemplary items are "I can spend hours on a single problem because I just can't rest without knowing the answer" and "Thinking about solutions of difficult conceptual problems can keep me up at night." The second, rather

experiential factor in Kashdan et al.'s (2020) multidimensional curiosity concept, *Joyous Exploration*, describes "the pleasurable experience of finding the world intriguing" (p. 1) and corresponds to the enthusiasm about learning new things—and thinking deeply. Two exemplary items for this factor are "I seek out situations where it is likely that I will have to think in depth about something" and "I enjoy learning about subjects that are unfamiliar to me." Despite distinguishing between two factors, their relatedness is evident from the similarity of the respective item wordings.

Students' scientific thinking

The Q-Assessment of Undergraduate Epistemology and Scientific Thinking (QUEST; Zagumny, 2018) measures "the dispositional attitudes toward scientific thinking and intellectual curiosity among undergraduate students" (p. 928). While differentiating an internal from an external aspect— general IC (e.g., "When learning about something new or experiencing something new, I often lose track of time") vs. school-specific IC (e.g., "I like learning new things even if I don't need them for school or my job")—the QUEST resembles age-appropriate epistemic curiosity.

Curiosity at work

Mussel et al. (2012) proposed yet another specific, work-related curiosity scale (WORCS), which employs a high proportion of IC-like items. Exemplars are "I enjoy pondering and thinking" for enjoying the effortful cognitive activity and "I keep thinking about a problem until I've solved it" for the mental restlessness before a task is completed or a problem solved.

Readiness to Learn

The *Readiness to Learn Scale* (RtL; Smith et al., 2015) measures adult "readiness to engage in learning activities" (p. 4) related to challenges, problems, or tasks. Like the previous inventories, the scale reflects a range of IC-like characteristics such as mental restlessness upon challenge (e.g., "I like to get to the bottom of difficult things"), striving to solve difficult problems to gain better understanding (e.g., "I like to figure out how different ideas fit together"), and engaging in learning as such (e.g., "I like learning new things").

In sum, many different but related conceptualizations exist, and considerable overlap in concepts and their measurement is evident. The application scenarios for IC and related concepts are manifold, but also similar. Here, we draw the following interim conclusion: The high relevance of IC has led to an abundance of scales. These diverse assessment approaches have substantial overlap in terms of scale correlations and shared item variance in factor analyses (Mussel, 2010; Woo et al., 2007). Specifically, Powell et al. (2016) ran a factor analysis across the items of three prominent scales (NfC, TIE, and ECS), concluding that up to six factors are necessary to explain all reliable item covariance. However, they suggested no new scales. Further, for IC to be a human universal, the measurement has yet to be rigorously compared across languages to rule out cultural

peculiarities (e.g., in the cultural stress of academic education or mere access to higher education facilities independent of socio-economic background). No psychometrically valid instrument for measuring IC cross-culturally has emerged so far. The ICS, developed for the context of the OECD's Programme for the International Assessment of Adult Competencies (PIAAC; Organisation for Economic Co-operation and Development (OECD), 2019), may fill this void.

Measurement of intellectual curiosity in the PIAAC context

The ICS was developed by a team of experts gathered for the PIAAC pilot studies on non-cognitive skills. PIAAC aims at comparing adult competencies across different countries, for the sake of assessing the human capital of participating countries accurately and comprehensively—and while ensuring international comparability (Rammstedt, 2013). The purpose of the PIAAC non-cognitive pilot study was to develop and test various non-cognitive scales that might be included in the main study. Besides the ICS, the piloted non-cognitive scales included, for example, Big Five scales.

The ICS was newly composed for the PIAAC non-cognitive pilot and consists of six items that OECD experts selected from existing inventories with a high chance of comparability across countries (see Table 1). Regarding their origin, items were taken from several instruments for a sufficiently broad concept representation. The items IC1–IC4 as used in the PIAAC survey were taken from the RtL scale (Gorges et al., 2017). IC1–IC3 were originally taken for PIAAC from the motivation scales of the widespread *Motivated Strategies for Learning Questionnaire* (MSQL; Duncan & McKeachie, 2005), and IC4 from the *Achievement Motivation Questionnaire* (Harackiewicz et al., 1997), frequently used in educational studies. IC5 and IC6 were taken from the PISA 2012 *Openness to Problem Solving Scale* (OECD, 2013). For all items, participants indicated the applicability of the statements to themselves on a five-point Likert-type scale (1 = *Not at all* to 5 = *To a very high extent*) in response to the question "To what extent [do] the following statements apply to you?"

The source version of the ICS was translated from English language to French, German, Japanese, Polish, and Spanish (for the final translations also of the response options, see Tables A1–A5). The translations were derived through a modified TRAPD approach (Harkness, 2003), which usually

Table 1. Intellectual Curiosity Scale with six items (ICS; English source version used for the USA).

Variable label	Item wording
IC1	I like learning new things.
IC2	I like to get to the bottom of difficult things.
IC3	I like to figure out how different ideas fit together.
IC4	If I don't understand something, I look for additional information to make it clear.
IC5	I seek explanations of things.
IC6	I like to solve complex problems.

Note. Introductory Question: "To what extent [do] the following statements apply to you?" Response options: 1 = *Not at all*, 2 = *Very little*, 3 = *To some extent*, 4 = *To a high extent*, 5 = *To a very high extent*. Variable labels reflect the order of item presentation.

comprises five steps: translation, review, adjudication, pre-testing, and documentation. In this case, after outsourcing the process from the OECD to a professional translation service, for each of the five languages two expert translators provided independent translations. These materials were then reviewed and adjudicated, after which psychometric experts who were native speakers of each target language provided additional feedback on the adjudicated items (an additional step beyond typical TRAPD stages). Before the ICS can be recommended, we investigate reliability and validity after inspecting measurement invariance across six languages besides gender, age, and education.

Materials and methods

Sample

The PIAAC international pilot studies on non-cognitive skills recruited participants from Germany, France, Japan, Poland, and Spain (data available from OECD, 2018b). Data collection took place in 2016–2017 (GESIS, 2021; Maehler & Rammstedt, 2020). Together with participants from the U.S. (data available from OECD, 2018a), we included 5,557 respondents who matched the quality-filtered sample described by Partsch and Danner (2021). There were no missing values on the items and scales of interest (and negligible missingness on a few socio-demographic variables). Table 2 shows the socio-demographics for all six country-specific samples. Mean age was 43.19 years ($SD = 12.70$). The analytical sample was rather balanced in terms of the gender distribution (54% identified as female; the rest all identified as male), though French (60%) and U.S. (59%) participants both tended toward uneven gender distributions. Further information about the instruments (including translated ICS item wordings) and the study design is accessible from the documentation (OECD, 2018a, 2018b) and from the Supplemental Online Materials (ICS_SOM.pdf and ICS_SOM_Invariance_Validity.xlsx; https://osf.io/dzfu3/).

Measurement instruments for the nomological net

We aimed to validate ICS by locating it in a broad set of individual-difference constructs that were available in the dataset. We selected these variables based on their conceptual closeness to IC. We focus on the variables for which we expected positive correlations with the ICS (convergent validity). As noted below, for a few other variables we expected to find lower (or near-zero) correlations, providing evidence for discriminant validity. Unless stated differently, the response scale for all variables was a 5-point Likert-type scale (ranging from *strongly disagree* to *strongly agree*).

Big Five Inventory-2

Following Kashdan et al. (2020), we correlated the ICS with each Big Five domain—*Open-Mindedness*, but also *Extraversion*, *Conscientiousness*, *Agreeableness*, and *Negative Emotionality (Emotional Stability)*—as measured with the Big Five Inventory-2 (BFI-2; Soto & John, 2017). Note that the OECD applied fully labeled instead of endpoint-labeled response options (*strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree*). Each domain comprises three narrow facets (four items each), which we also correlated with the ICS. We expected the ICS to show convergent associations with the Big Five domain *Open-Mindedness* as measured with the BFI-2, most strongly so with the *Intellectual Curiosity facet*.[1] At the same time, we expected slightly weaker relationships with the other two *Open-Mindedness* facets (i.e., *Aesthetic Sensitivity* and *Creative Imagination*). Yet, lower correlations should particularly emerge for the other four (orthogonal) personality domains (C, E, A, N). In conjunction with the evidence for convergent validity, such a pattern would nicely demonstrate discriminant validity of the ICS.[2]

Sensation Seeking and Perseverance

The PIAAC pilot asked questions about *Sensation Seeking* (Whiteside & Lynam, 2001) and *Perseverance* (OECD, 2018a, 2018b), each construct being measured with five items and having some bearing on IC. In the OECD context, both aspects are considered components of self-control (besides *Negative Urgency* and *Premeditation*). Exemplary items are "I quite enjoy taking risks" (*Sensation Seeking*) and "I continue working on tasks until everything is perfect" (*Perseverance*). While *Sensation Seeking* corresponds to being stimulated by unfamiliar environmental stimuli (e.g., thrills, tasks, problems, or simply information), *Perseverance* describes the eponymous notion to fill knowledge gaps or

Table 2. Socio-demographic sample composition in six OECD countries.

	France	Germany	Japan	Poland	Spain	USA
Participants						
Male	370	549	367	427	466	361
Female	565	569	411	457	497	518
Total	935	1118	778	884	963	879
Age						
Minimum	16	18	18	18	18	16
Maximum	65	65	65	65	65	65
Median	44	48	46	41	42	43
M	42.91	45.21	44.67	41.37	42.09	42.63
SD	13.07	12.25	12.36	13.20	11.80	13.12
Education						
Primary or High school	513	598	297	390	240	208
College/Vocational training	312	142	153	99	277	345
Tertiary education	110	378	328	395	446	326

Note. Primary school was joined to high school due to the low number of participants who only obtained basic schooling (between 0 and 29 in the countries).

---

[1]For each construct, Table 5 provides reliability estimates based on factor loadings (McDonald's Omega) taken from separate CFA models, usually achieving acceptable model fit. For each BFI-2 domain, unidimensional and bifactor models attained poor fit, so no Omega values can be reported. As a crude reliability estimate, we report Cronbach's Alpha without adopting its assumption of "essential tau-equivalence."

[2]To establish discriminant validity, we screened criteria available from the PIAAC pilot dataset. We considered *Traditionalism* (assessed with eight items, e.g., "I support long-established rules and traditions."), because the Schwartz value *Tradition* had previously shown no overlap with IC, so we expected hardly any relevant association between IC and *Traditionalism* (see Grüning & Lechner, 2023; Kashdan et al., 2020). Yet, *Traditionalism* is a less than optimal representation of *Tradition*, so we refrain from discussing these findings; instead, we refer the reader to Table 5 and the country-specific Tables D1–D6 in ICS_SOM.pdf. Similarly, we deemed *Social Trust* conceptually distinct from IC. However, a mere two PIAAC-specific items were the only basis to test the correlation, which we had suspected to be in the vicinity of zero.

think about problems until solved. We expected *Perseverance* to show a medium-sized correlation with the ICS, as the criterion not only reflects persistence but also the strong desire for ultimately solving problems. Yet, the tendency to enjoy cognitive challenges—*Sensation Seeking*—relates to the joy of experiencing novel stimuli (also embedded in the ICS). It reflects more the experiential side of curiosity than it reflects intellectual needs. A positive correlation can still be expected based on the desire to gain information to reduce the unknown.

Job Orientations

Within PIAAC, the Job Orientations Scale (JO) illuminated *Learning Opportunities*, which should clearly correspond to the ICS (OECD, 2018a, 2018b). The concept was represented by two items ("A job that allows you to learn new skills" and "A job that offers good training opportunities"), each rated on a 5-point Likert-type scale (ranging from *not at all important* to *very important*).

Inquisitive vocational interests

Five items (e.g., "Develop a way to better predict the weather" and "Study ways to reduce water pollution") formed an index of respondents' inquisitive interests. Participants expressed their vocational interests on a 5-point Likert-type scale (ranging from s*trongly dislike* to *strongly like*). Compared to JO, the ICS scores should correspond moderately—but not strongly—to job-related *Inquisitive Interests*, due to manifold influences during the formation of vocational interests and a rather specific measurement focus on natural sciences (medication, pollution, chemical experiments).

Statistical analyses for scale validation

After inspecting descriptive statistics of ICS item and scale scores, we evaluated the ICS's potential for a future standard measure of IC. We first checked unidimensionality by exploratory and confirmatory factor-analytical approaches. We established an acceptable measurement model, which served as the basis for testing of measurement invariance (MI) across different grouping variables (gender, age, education, and language/countries). Only after the crucial question of comparability has been answered can one inspect the psychometric criteria of reliability and construct validity, allowing for a comparison of the country results. Having established the psychometric viability of the ICS, we finally compared latent mean differences to demonstrate the utility of the ICS for substantive research.

The following analyses were originally conducted with R version 4.0.3 (R Core Team, 2020) and the following R-packages: afex (Singmann et al., 2021), dplyr (Wickham et al., 2021), EFAtools (Steiner & Grieder, 2020), emmeans (Lenth, 2021), ggplot2 (Wickham, 2016), graphics (R Core Team, 2020), lavaan (Rosseel, 2012), MVN (Korkmaz et al., 2014), nortest (Gross & Ligges, 2015), pastecs (Grosjean & Ibanez, 2018), psych (Revelle, 2020), QuantPsyc (Fletcher, 2012), and utils (R Core Team, 2020). All R-scripts are available from the first author upon request.

Factorial validity: Establishing the dimensionality and CFA measurement model

To establish for each country the intended unidimensional factor structure for the ICS, which assumes that all items load on a single common factor, we drew on Velicer et al.'s (2000) revised MAP-test, Horn's (1965) parallel analysis (PA; once run with principal components and once run with principal axis factoring). After it became clear that the assumption of strict unidimensionality had to be relaxed, we utilized the index of proximity to unidimensionality (IPU; Raykov & Bluemke, 2021) to evaluate if essential unidimensionality was tenable, before we turned to modification indices and fit indices for comparing strictly and essentially unidimensional CFA measurement models.

We used the robust Maximum Likelihood (MLR) estimator (with Huber-White correction of standard errors and a Yuan-Bentler equivalent test statistic) to compensate for non-normal distributions of the ordinal data (Buchholz & Hartig, 2020; Marsh et al., 2018). For identification of the basic measurement model, we fixed the latent factor variance to 1 (and the latent factor mean to 0). Given that χ^2-tests are sensitive to small deviations in large samples (Bentler & Bonett, 1980; Fischer & Karl, 2019), we report them for descriptive purposes. We prefer to use goodness-of-fit indices to evaluate model fit (Chen, 2007, 2008; Fischer & Karl, 2019; Svetina et al., 2020): A unidimensional factor structure is supported when all six items load on the same factor (λs > .40 while conventional criteria indicate adequate model fit, such as Comparative Fit Index (CFI) > .90, Root Mean Square Error of Approximation (RMSEA) < .08, and Standardized Root Mean Square Residual (SRMR) < .08. Very good model fit would follow from CFI ≥ .95, RMSEA ≤ .05, and SRMR ≤ .05, though we caution against using cutoffs rigidly as they only apply to models that match the simulation from which they derive (Hu & Bentler, 1999).

We first investigated the measurement model via single-group CFA models for each country. Only when this basic measurement model is also comparable across groups can one proceed to a multiple-group CFA (Byrne, 2008; Cieciuch & Davidov, 2015). However, good model fit may only be attainable after dealing with the violations of strict unidimensionality: Poor fit may hint at residual covariances (i.e., error correlations), and additional parameters may then be needed for MI testing.

Measurement invariance testing with multiple-group confirmatory factor analysis

Multiple-group CFA (MG-CFA) was applied across the six countries to test four increasingly restricted MI levels—configural, metric, scalar, and residual invariance—by imposing more and more parameter equality constraints (Chen, 2008; Cieciuch & Davidov, 2015; Davidov et al., 2018; Fischer & Karl, 2019; Milfont & Fischer, 2010; Vandenberg & Lance, 2000). We checked cross-cultural MI both for the ICS and the constructs used for its validation, and for the ICS we ran further MI checks across gender, age, and formal education groups.

At the configural level, the MG-CFA model does not require any parameter equality constraints, merely an identical item-factor configuration (including the presence or

absence of residual covariances). At the next level, metric MI imposes equal parameters (unstandardized factor loadings) across groups. Only metric MI allows the meaningful comparison of variances and covariances as per correlation or regression analyses. For scalar MI, the item intercepts are additionally fixed to equality across groups, which allow meaningful mean-level comparisons across groups. If one were interested in using and comparing manifest item and scale scores, the strictest MI level can be tested by additionally imposing equal residual item variances, which, if it holds, demonstrates equal measurement error in each group (Cheung & Lau, 2012). If the model shows insufficient fit at a specific MI level, researchers often strive for *partial* invariance. In this case, they may free parameters for some non-invariant items while retaining equal loadings, intercepts, and/or residual parameters for the invariant items. In many application scenarios, achieving partial MI is sufficient for legitimate group comparisons (Borsboom, 2006; Byrne et al., 1989; Steenkamp & Baumgartner, 1998).[3]

We first evaluated configural MI (in the way how we evaluated overall fit of the single-group CFA measurement model in each country). Then, we determined whether a more parsimonious MI level held—or whether a less stringent MI model was needed—on the basis of delta-fit heuristics that may indicate a loss of model performance (ΔCFI $\leq$ .01; ΔRMSEA $\leq$ .015; ΔSRMR $\leq$ .03 or $\leq$ .01 for metric or scalar invariance, respectively). We sought convergence with the Bayesian Information Criterion (BIC) as it does not require arbitrary cutoff heuristics but compares two models against each other (lower BIC values indicate a better parsimony-accuracy tradeoff; Byrne, 2016).

Reliability: Composite reliability

Whereas latent variable models correct for measurement error, researchers often use manifest scale scores that are subject to measurement error. We estimated scale reliability with McDonald's omega (ω). While Cronbach's α simply assumes equal factor loadings across all items, ω uses the empirical loadings from an acceptable unidimensional CFA model (McNeish et al., 2018). It shows the percentage of variance in scale scores that is true score variance explained by the latent variable.

Construct validity: Manifest and latent validity correlation coefficients

To assess convergent and discriminant validity, after computing scale scores as proxies by averaging the (unit-weighted) item responses, we calculated the Pearson correlation coefficients between ICS and validation criteria. We also estimated the structural correlations between latent variables

in multiple-group structural equation models (MG-SEM; Noar, 2003). If possible, we used a multistage approach that first tested MI for each criterion variable across countries (like for the ICS). In a second step, we combined the two models fixing the parameters to those obtained from single-construct models before investigating convergent validity (e.g., with BFI-2's *Open-Mindedness scale*) and discriminant validity (e.g., with *Social Trust*). We used this two-step procedure rather than the simultaneous modeling of paths to estimate validity at the construct level after correcting for measurement error while avoiding interpretational confounding (cf. McNeish & Wolf, 2020).

Sensitivity to between-group differences

The comparison of group means requires at least partial scalar invariance. Therefore, we investigated whether the ICS can legitimately differentiate gender, age, education, and language groups. We did not form any specific hypotheses about gender differences in IC due to conflicting findings on gender effects (Cacioppo et al., 1996; Engelhard & Monsaas, 1988; Powell et al., 2016). However, we hypothesized that IC (a) declines with advancing age (Chu et al., 2021; Dellenbach et al., 2008; Engelhard & Monsaas, 1988; Zimprich et al., 2009) and (b) increases with higher formal education levels attained (Orcutt & Dringus, 2017; von Stumm et al., 2011; von Stumm & Ackerman, 2013). The scarcity of literature on cross-national comparisons in IC rendered our cross-country analysis exploratory in nature, providing a check of sufficient sensitivity to group differences.

Results

Descriptive statistics

For a descriptive analysis of the six ICS items and their intercorrelations, see Tables B1 and C1 as part of the Supplemental Online Material (ICS_SOM.pdf). For country-specific tables, see Tables B2 and C2. The descriptives showed that item responses were not identically distributed across countries (the utilized ranges differed per item and country). To not discard any information during CFA modeling, exploiting all available information required a maximum likelihood approach with assumptions about normality—a method we preferred over collapsing response categories to arrive at the same number of ordinal categories (before any categorical estimator might be applied to test a measurement model; see Table B2). We inspected, for each country, univariate normality of the ICS items and scale scores with the Shapiro-Wilk test and the Shapiro-Francia test (Royston, 1983; Shapiro & Wilk, 1965). As the statistics cannot be computed for sample sizes with $N > 5000$ (Royston, 1995), we tested normality at the country level. All the test statistics yielded nearly identical figures with *p*-values < .001, so that non-normal distributions must be assumed (see Table B3). Consistent with this pattern, the Henze-Zirkler and Mardia tests of multivariate normality also failed (*p* < .001; Henze & Zirkler, 1990; Mardia, 1970), necessitating the use of *robust* ML (MLR) estimation to handle nonnormal

---

[3]For identifying MG-CFA models, we used the identification by reference-group approach. As the source language was English (OECD, 2018a), U.S. participants served as the reference group. The approach requires constraining (and freeing) the latent variances and latent means (Schroeders & Gnambs, 2018): At the configural level, in each group the variance and mean of a latent factor are set to unity and zero, respectively. At the metric level, the variance is freed in all but the reference group. At the scalar level, additionally the means of all but the reference group are freed. At the residual level, no additional identification constraints are required.

skew and kurtosis (for descriptive statistics of ICS scale scores, see Table B4).

Factorial validity

The MAP test and PCA-based Parallel Analysis suggested a single strong dimension in each country (R^2 = 62–69%). Except for Japan, the principal-axis based PA consistently suggested that a secondary dimension is needed to explain a small part of the common variance above chance level (see Figure 1 and Table C3). The loading coefficients in two-dimensional exploratory factor analyses consistently identified the item-pair IC4 & IC5 as the driver of this local dependency (see also Table C4). Using Raykov and Bluemke's (2021) recent CFA-based "index of proximity to unidimensionality" (IPU), we quantified the deviation from unidimensionality by computing the variance proportions attributable to the general factor of interest (π_G) vs. the local factor for the item-pair (π_L) besides residual uniqueness factors (π_E; Table C3 for IPU estimates). In the absence of universal guidelines for interpreting IPU, we report IPU to encourage researchers to gather more experience with it. The initially suggested—rough—guideline (relative proportions of 70:20:10) was too strict, as the general factor remained below the 70%-threshold though its 10-fold dominance over the local factor was evident (π_G = 55%–63% vs. π_L = 3.2%–6.2%).

Turning to the model fit of the unidimensional model, we drew on the *robust* model fit and *robust* fit indices in lavaan's MLR output for CFA models. In each country the factor loadings were strong (see upper part of Table 3). There was good fit according to CFI and SRMR, although RMSEA did not pass the conventional threshold for acceptable model fit in all the groups (RMSEA > .08). Modification indices consistently pointed out that adding a parameter for the non-negligible residual covariance (IC4–IC5) could improve model fit significantly (and would be more informative than any other model adjustment; see Table C4). An exception was Japan, where this residual covariance merely ranked second place, following closely behind the residual covariance IC2–IC6 (estimated χ^2-improvements were 33.53 vs. 26.93).[4] Looking at the expected parameter change (EPC-standardized), the size of the IC4–IC5 error correlation indicated a minor effect in all countries, confirming the impression conveyed by IPU. We attributed the necessity for an additional covariance to the semantic closeness of "seeking explanation of things" (IC5) and "looking for additional information" when not comprehending a matter (IC4). Though not a wording effect proper, this similarity can drive a method-factor beyond the variance explained by the general common factor, though it might also reflect a substantive facet in larger (more redundant) item pools.

Regarding MI testing, the need for an additional equality constraint for this residual covariance depends on whether one considers it a part of the theoretical measurement model (an a priori facet) or a post hoc modification to explain a minor amount of unexpected wording covariance, independent from the intended measurement of the construct of interest. While the definition of a facetted measurement approach would necessitate an equality constraint that requests strictly the same amount of secondary item covariance across all countries, a post hoc adjustment may also allow to freely estimate the error covariance across countries (Avery et al., 2007; Byrne, 2008; Byrne & van de Vijver, 2017). Enforcing a strictly unidimensional model tended to overestimate the factor loadings except for the factor loadings of items IC4 and IC5 which tended toward underestimation. In the end, we accepted the essentially unidimensional measurement model with a freely estimated residual covariance for IC4–IC5.[5] In each country—including Japan— these items defined the secondary factor as per their loadings. Including their correlation consistently yielded favorable model fit (see lower part of Table 3 for country-specific fit indices and factor loadings). Consequently, we used this adjusted model for all measurement invariance testing and for the assessment of construct validity.

Measurement invariance

We first fit to each analyzed group the single-group CFA model, before using multiple-group CFA for testing MI across gender, age, education, and countries (languages) as grouping variables, while pooling across the non-focal grouping variables (see Table 4). We also ran cross-checks within countries to ascertain the MI level accepted for gender, age, and education was consistently attainable.

Gender

The essentially unidimensional model (including IC4–IC5) achieved good model fit in each gender group. The "worst" fit resulted for participants identifying as male, $\chi^2(8)$ = 47.71, CFI = .993, RMSEA = .054, SRMR = .014, which supports an excellent psychometric model. Combining both gender groups into an MG-CFA model, with model parameters estimated freely (including the error correlation IC4–IC5), resulted in equally good fit, supporting configural MI. Restricting each item's factor loading to equality across genders resulted in as good model fit, hence metric MI was clearly tenable (ΔCFI = −.001, ΔRMSEA = −.006, ΔSRMR = +.005, ΔBIC = −32.53). While the fit indices suggested the tenability of scalar MI (ΔCFI = −.003, ΔRMSEA = +.004, ΔSRMR = +.005), with a ΔBIC value of +21.35 the metric MI model appeared to be preferable. However, compared to the configural model, BIC still favored scalar

---

[4]The weaker Japanese residual covariance for IC4–IC5 than for IC2–IC6 might be due to the complexity of the Japanese language regarding word families. Due to the variety of characters, and the different combinations thereof, there are more nuanced versions of "to understand" and "to explain" (M. Wierzba, personal communication, August 10, 2021; C. Stoica, personal communication, August 23, 2021). A popular Japanese dictionary explicates "to explain" with the word for "to understand" as used in the ICS (第2版, 2021). Including IC4-IC5 improves fit and acknowledges the semantic relation between the Japanese item wordings.

[5]We note here that the same invariance levels across countries were achieved with consistency checks that set an additional equality constraint for the residual covariance IC4–IC5 across countries (as if the residual covariance or the facet it represents would have been part of the theoretical measurement model).
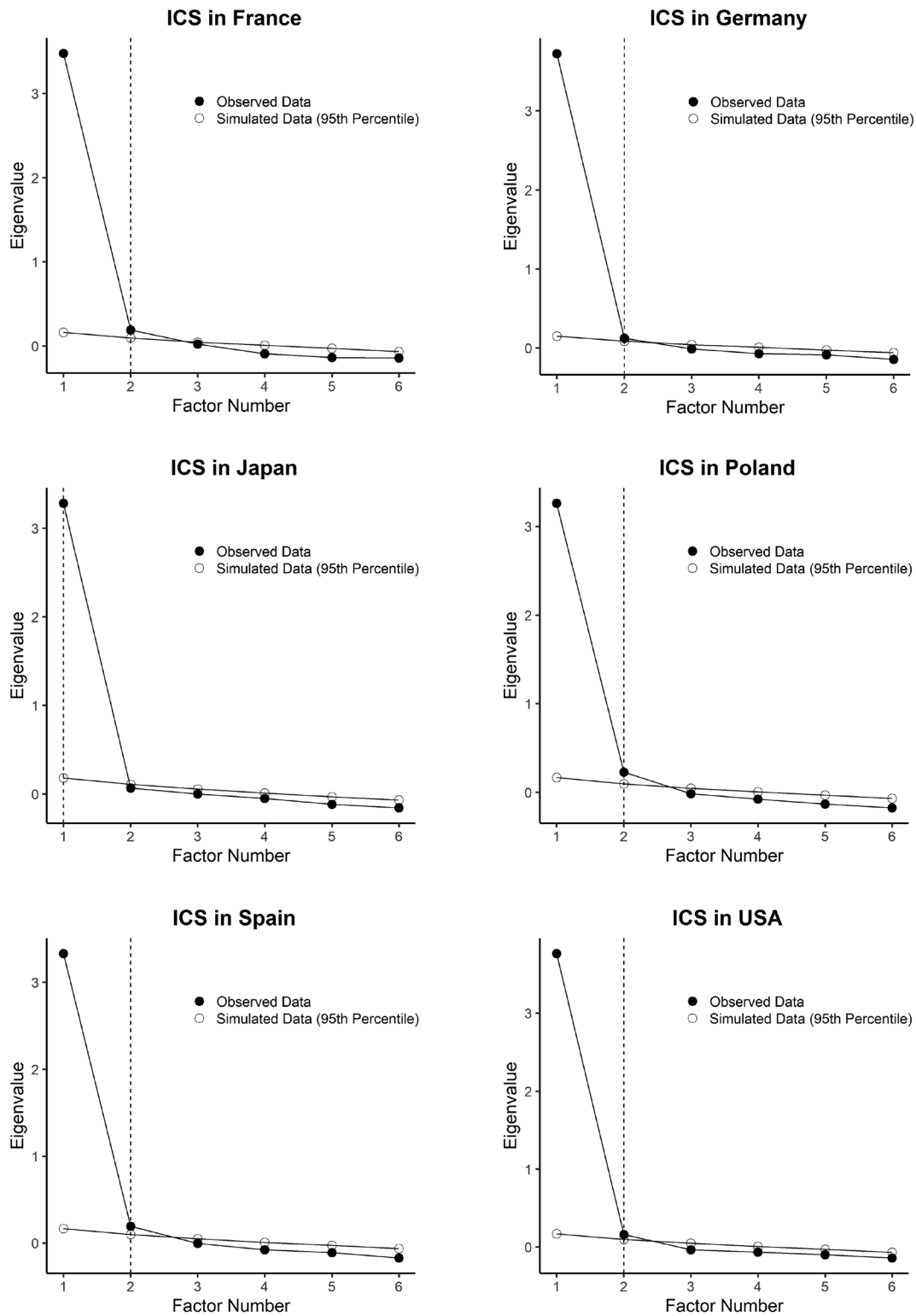
Figure 1. ICS scree plot and parallel analysis (based on principal axis factoring) per country.

invariance.[6] Also, the strict MI model (after adding equal residual variances) fitted as well as the scalar MI model (ΔCFI = 0, ΔRMSEA $=-.003$, ΔSRMR = 0, ΔBIC = -22.83), and in terms of BIC the strict model performed as well as the metric MI model. Considering the overall fit in conjunction with Chen's (2007) delta-fit guidelines, we consider the ICS strictly invariant across gender groups. We replicated the tenability of strict invariance when testing MI separately in each of the six countries (see ICS_SOM_Invariance_Validity.xlsx).

---

[6]A robustness check showed that freeing a single intercept (IC6) was sufficient to attain partial scalar invariance (CFI = .992, RMSEA = .047, SRMR = .019). This time BIC = 66,212 supported the model over the metric invariance model, yet the intercept difference across gender groups was small ($\Delta\tau$ = 0.11).

Table 3. Model fit and factor loadings of tested ICS measurement models (per country).

	χ^2	CFI	RMSEA	SRMR	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
Unidimensional model ($df=9$)										
France	144.31***	.933	.156	.043	.71	.80	.78	.77	.81	.72
Germany	111.76***	.965	.121	.032	.68	.86	.85	.73	.78	.83
Japan	36.24***	.978	.083	.028	.65	.87	.76	.75	.65	.77
Poland	131.98***	.931	.148	.048	.70	.83	.72	.68	.74	.78
Spain	146.74***	.940	.141	.043	.72	.84	.71	.70	.73	.79
USA	107.79***	.958	.132	.036	.77	.82	.87	.77	.75	.79
Essentially unidimensional model with IC4–IC5 ($df=8$)										
France	40.15***	.985	.079	.025	.73	.83	.80	.70	.74	.72
Germany	31.65***	.992	.060	.015	.69	.87	.85	.70	.75	.83
Japan	21.85**	.989	.062	.021	.65	.88	.76	.73	.62	.78
Poland	31.98***	.987	.069	.022	.71	.85	.73	.62	.69	.78
Spain	35.40***	.988	.067	.020	.72	.86	.72	.65	.68	.79
USA	22.43**	.994	.053	.017	.77	.83	.88	.73	.71	.80

Note. N_{total} = 5,557. χ^2 = scaled χ^2 test statistic, df = Degrees of Freedom, CFI = (Robust) Comparative Fit Index, RMSEA = (Robust) Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, λ_{1-6} = Standardized Factor Loadings (Item 1–6).
All factor loadings are significant at $p < .001$, other p-values: **$p < .01$, ***$p < .001$.

Age

Before comparing the three age groups (16–29 yrs., 30–49 yrs., and 50–65 yrs.), single-group CFA models confirmed good model fit. The "worst" fit statistics resulted for the young cohort (χ^2 = 24.07, CFI = .990, RMSEA = .059, SRMR = .020), while overall indicating excellent model fit. Regarding MG-CFA models, both configural and metric MI held (ΔCFI = .000, ΔRMSEA = −.008, ΔSRMR = +.005, ΔBIC = −71.88). As with gender, scalar MI was tenable across age groups according to Chen's delta-fit criteria (ΔCFI = −.005, ΔRMSEA = +.009, ΔSRMR = +.006), though BIC cautioned against it (ΔBIC = +33.22). We again sided with Chen's cutoff criteria and accepted scalar MI.[7] However, a similar discrepancy arose when testing strict MI. This time, the increase in misfit was tangible (largest χ^2-increase so far), and this time, despite high model parsimony, BIC increased further rather than returning to preferable lower levels (as was the case for gender). Still, based on Chen's criteria, strict MI appeared tenable (ΔCFI = −.009, ΔRMSEA = +.010, ΔSRMR = +.004, ΔBIC = +91.08), and was confirmed by the high levels of overall model fit.

Testing MI separately in countries clarified the muddy waters. Based on BIC, which is straight forward in single-group analyses, strict MI was tenable in all countries except Poland, for which only scalar MI held (see ICS_SOM_Invariance_Validity.xlsx). (By contrast, based on Chen's criteria, Poland attained only partial scalar MI (free intercept for IC6: τ_{young} = 3.516, $\tau_{medium-aged}$ = 3.698, τ_{old} = 3.768), while Japan attained scalar MI.) Overall, the comparability of age groups was supported, though the Polish sample fell short, and this was reflected in the unfavorable BIC values in a simultaneous analysis of all countries. If we had to speculate, three potential explanations for the Polish

[7]When we tested a partial invariance model with a single free intercept (IC1), all fit indices (CFI = .992, RMSEA = .045, SRMR = .019) and BIC = 66,091 agreed on its tenability. It should be noted that the maximum absolute intercept difference resulting across the three age groups was rather small ($\Delta\tau \leq 0.21$).

discrepancy come to mind. An innocuous explanation blames mere sampling error. Alternatively, the Polish item translation for IC6 may have been suboptimal, unintentionally undermining the fairness for the age cohorts. A substantive explanation, however, cannot be ruled out: National specificities of the economy may have affected the age groups in Poland differently than in other countries. The historical rift of the job market (turning a former socialist economy into a free labor market) may have altered the relevance of the content surveyed in IC6 for the job market—hence the fluctuating item difficulty for the age cohorts in the context of the PIAAC non-cognitive pilot.

Education

Comparing participants with low, intermediate, and high levels of formal education, the single-group CFA models provided good model fit in all groups (the "worst" fit resulted for participants in the low education group, χ^2 = 48.43, CFI = .993, RMSEA = .058, except for the "worst" SRMR = .016 that emerged for intermediate education levels). Using the same hierarchical procedure as with gender and age groups before, we evaluated MI for the education groups. Increasing the number of parameter equality constraints hardly decreased model fit. Not only configural MI, but also metric MI (ΔCFI = −.001, ΔRMSEA = −.010, ΔSRMR = +.004, ΔBIC = −109.61), scalar MI (ΔCFI = −.001, ΔRMSEA = −.003, ΔSRMR = +.003, ΔBIC = −90.97), and strict MI (ΔCFI = −.001, ΔRMSEA = −.002, ΔSRMR = −.001, ΔBIC = −105.53) held. This time, BIC clearly supported the strict MI model, corroborated within each country (see ICS_SOM_Invariance_Validity.xlsx).

Countries

Single-group CFA models for countries had already suggested good model fit when establishing the essentially unidimensional measurement model (see Table 3; the "worst" fit indices emerged for France: χ^2 = 40.15: CFI = .985, RMSEA = .079, SRMR = .025). Evaluating MI across countries is crucial, because it concerns the utility of the scale for the purpose for which it was invented: cross-national comparisons. At the same time, it is a very rigorous test of equivalent functioning of the scale across six countries despite language differences, national adaptations, and cultural specifics that may pertain to IC. Comparing the metric to the configural MI model suggested that metric MI was tenable (ΔCFI = −.004, ΔRMSEA = −.003, ΔSRMR = +.023, ΔBIC = −109.16). By contrast, testing for scalar MI decreased model fit considerably, as is often the case in multinational studies (ΔCFI = −.029, ΔRMSEA = +.031, ΔSRMR = +.021, ΔBIC = +341.34).

Hence, we approached partial scalar MI. Freeing the first equality constraint implied by the modification indices (IC1 intercept) still resulted in a substantial decrease of fit from metric MI (ΔCFI = −.018, ΔRMSEA = +.021, ΔSRMR = +.013, ΔBIC = +184.22). When freeing another intercept (IC3) the model fit improved, yet the improvement was ambiguous given that CFI slightly exceeded the threshold for accepting scalar MI and given that BIC favored metric

Table 4. ICS measurement invariance models for grouping variables.

	χ^2 (df)	$\Delta\chi^2$	CFI	RMSEA	SRMR	BIC
Gender						
Configural	100.53 (16)	–	.994	.054	.012	66,255
Metric	114.66 (21)	9.42	.993	.048	.017	66,223
Scalar	169.72 (26)	64.88***	.990	.052	.022	66,244
Strict	185.94 (32)	18.67**	.990	.049	.022	66,221
Age						
Configural	101.73 (24)	–	.994	.052	.012	66,184
Metric	118.76 (34)	12.06	.994	.044	.017	66,112
Scalar	216.84 (44)	117.79***	.989	.053	.023	66,145
Strict	352.72 (56)	128.62***	.980	.063	.027	66,236
Education						
Configural	116.98 (24)	–	.993	.056	.013	66,306
Metric	130.58 (34)	5.10	.993	.046	.015	66,225
Scalar	167.30 (44)	36.21***	.992	.044	.018	66,175
Strict	199.74 (56)	34.02***	.990	.043	.020	66,123
Countries (Languages)						
Configural MI	181.41 (48)	–	.989	.066	.017	64,637
Metric MI	282.03 (73)	102.15***	.985	.063	.040	64,528
Scalar MI	755.66 (98)	561.18***	.956	.094	.061	64,869
Partial scalar (τ_1)	585.23 (93)	356.80***	.967	.084	.053	64,712
Partial scalar (τ_1, τ_3)	461.85 (88)	211.46***	.974	.076	.047	64,608
Partial scalar (τ_1, τ_3, τ_2)	366.54 (83)	98.34***	.980	.068	.043	64,537

Note. $N = 5,557$. χ^2 = scaled χ^2 test statistic (all $p < .001$), $\Delta\chi^2$ = change of the (Satorra-Bentler corrected) χ^2 test statistic, df = Degrees of Freedom, CFI = (Robust) Comparative Fit Index, RMSEA = (Robust) Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, BIC = Bayesian Information Criterion.
**$p < .01$, ***$p < 001$.

MI too (ΔCFI = −.011, ΔRMSEA = +.013, ΔSRMR = +.007, ΔBIC = +80.32). For an unambiguously partial invariance model, we freed a third intercept (IC2), which yielded acceptable fit (ΔCFI = −.005, ΔRMSEA = +.005, ΔSRMR = +.003, ΔBIC = +9.13). BIC indicated cautiousness about partial scalar MI compared to metric MI (yet clearly favored partial scalar over configural MI). We accepted the model with three unconstrained intercepts.[8]

Reliability

The partial scalar MI model is legitimate for estimating the reliability of ICS composite scores in each country. In the pooled sample, McDonald's ω amounted to .91, while the country-specific values varied slightly, albeit at excellent levels for a six-item scale, with ω_{France} = .89, $\omega_{Germany}$ = .91, ω_{Japan} = .88, ω_{Poland} = .88, ω_{Spain} = .88, and ω_{USA} = .91 (McNeish et al., 2018).

Construct validity

We established ICS validity for the validation constructs with the help of two different approaches: manifest and latent bivariate correlations. Whereas the manifest approach reflects the validity expected in some diagnostic settings with conventional

[8]A partial invariance model with a fourth freely estimated intercept (IC5) fitted significantly better according to χ^2, improving the fit indices further (CFI = .984, RMSEA = .064, SRMR = .041). Out of all MI models tested, this least restrictive partial invariance model would finally be adopted by BIC = 64,512. Note that the cross-country differences between freely estimated IC5 intercepts amounted to absolute $\Delta\tau \leq 0.17$, whereas the intercept ranges for IC1, IC2, and IC3 were roughly twice as large, $\Delta\tau \leq 0.37, 0.26, 0.33$, respectively. In many applications, it may hardly matter whether the fifth item intercept is treated as equal or varying.

Table 5. Reliability estimates and construct validity: Manifest and latent correlations (pooled sample).

Validation criterion	Intellectual curiosity scale			
	Reliability		Correlation	
Construct	α	ω	Manifest	Latent
Open-Mindedness (Domain)[a]	.85	–	.59	–
Intellectual Curiosity	.61	.61	.58	.76
Creative Imagination	.79	.77	.54	.63
Aesthetic Sensitivity	.80	.82	.36	.41
Extraversion (Domain)[a]	.86	–	.48	–
Energy Level	.71	.74	.48	.55
Assertiveness	.76	.77	.44	.53
Sociability	.78	.78	.30	.35
Conscientiousness (Domain)[a]	.89	–	.39	–
Productiveness	.77	.76	.42	.47
Responsibility	.68	.65	.36	.41
Organization[b]	.83	.84	.26	.30
Agreeableness (Domain)[a]	.81	–	.34	–
Compassion	.66	.67	.31	.42
Respectfulness	.69	.68	.30	.34
Trust	.58	.60	.22	.33
Negative Emotionality (Domain)[a]	.90	–	−.27	–
Depression	.81	.78	−.29	−.39
Emotional Volatility	.77	.77	−.25	−.30
Anxiety	.78	.78	−.18	−.22
Self-control				
Perseverance	.76	.76	.57	.68
Sensation Seeking	.80	.80	.39	.49
Job orientations				
Learning Opportunities	.74	.74	.53	.65
Vocational Interests				
Inquisitive	.83	.82	.37	.40
Other				
Traditionalism	.72	.74	.16	.21
Social Trust	.62	.62	−.05	−.07

Note. α = Cronbach's Alpha; ω = McDonald's Omega, with the two-item scales *Social Trust* and *Learning Opportunities (JO)* requiring equal loadings for model identification.
[a]Bifactor model could not be adequately fitted.
[b]Criterion MG-CFA measurement achieves configural MI only (we used partial metric MI with free loadings for BFI-2 items #3 and #33 for the MG-CFA, though in the pooled sample the countries must be considered metrically invariant).
All Pearson correlation coefficients are significant at $p < .001$ (two-tailed).

scale use (unit-weighted indexes), any associations between latent variables in SEM reflect true structural relationships (controlling for item unreliability). Latent correlations between ICS and other constructs resulted from construct-specific multiple-group models with at least metric invariance assumptions. Thus, we first tested metric MI via MG-CFA models (yet any two-item measures such as *Inquisitive Vocational Interests* require an additional equal loading constraint for model identification, which prevents proper MI testing). Then, we united the ICS measurement model with each construct and estimated the latent correlation.

For simplicity, Table 5 presents the correlation coefficients resulting in the pooled sample (for a country-specific comparison, see Tables D1–D6). Here, we only highlight country-specific findings if they deviate considerably in any of the countries. Beginning with basic personality variables, at the domain level we found the predicted highest correlation of IC with *Open-Mindedness*, likewise with the IC facet at the BFI-2's facet level. The correlation between the ICS and the BFI-2's IC-facet is also the strongest correlation we obtained, as these two are the theoretically most closely related scales. While being a clear confirmation of the ICS's convergent validity, the two constructs cannot be considered

identical, because the (disattenuated) structural correlation coefficients were far from unity. The pattern differed in Poland, where the highest correlation emerged with *Perseverance* rather than BFI-2's IC facet (manifest-r = .56 vs .49), even slightly so after correcting for measurement error (latent-r = .75 vs. .73). The strong correlation with *Perseverance* discounts the possibility that the Polish pattern is merely due to a weakness of the Polish BFI-2 (or of the ICS for that matter). Future research must determine if it is a real phenomenon that IC relationships in Poland deviate from those in other countries, or if a mere flaw in the translation process involuntarily incorporated nuances of perseverance into the Polish ICS wordings.

The other BFI-2 domains and facets supported discriminant validity for the ICS. Note that the correlation with *Extraversion* ranked second, supporting the relevance of IC during social encounters. However, a closer look reveals that it is *Energy Level* (but also *Assertiveness,* which is related to leadership skills) that drives the correlation with *Extraversion* foremost, and *Sociability* less so. Also note that, within the *Conscientiousness* domain, it was usually the BFI-2's *Productiveness* facet that correlated strongest with the ICS, which corroborates the relevance of IC for job-related performance. Among the basic personality domains, *Agreeableness* and *(Negative) Emotionality*, including their underlying facets, correlated least with IC.

Turning to the rather IC-specific validation constructs, IC correlated substantially with *Perseverance*, and to a lower extent with *Sensation Seeking*. This pattern speaks to the idea of tenacity being more relevant for IC compared to novelty-driven stimulus-search. IC correlated somewhat lower with *Job Orientations: Learning Opportunities* than with *Perseverance*. Further, being more distant in nature, *Inquisitive Vocational Interests*—with its highly selective focus on inquisitiveness about the natural sciences—typically correlated to an even lower extent.[9] As expected, but not discussed in detail here, the ICS usually obtained the lowest correlations with constructs we had chosen specifically for demonstrating discriminant validity (*Traditionalism* and *Social Trust*). Overall, the demonstrated sensitivity and specificity of the ICS for intellectual curiosity in the nomological net was sufficient.

Sensitivity to (known-)group differences
To depict the sensitivity of the ICS to group differences without being unfair toward socio-demographic subgroups, we compared the latent group means resulting from the MG-CFA models. Although overall statistically significant by conventional standards (Gonzalez & Griffin, 2001; Wald, 1943), the gender difference with a slight disadvantage for women was hardly noticeable (Δ = −0.07, p = .02). We obtained similarly negligible differences in Germany, Japan, and USA—and no significant differences at all in France, Spain, and Poland.

Using the medium age group as a reference, younger participants had significantly higher IC levels (Δ = +0.13, p = .002) and older participants significantly lower ones (Δ = −0.08, p = .01). The small differences became only tangible for the young–old contrast. Though the hypothesized age trend emerged quite clearly when pooling across all countries, the country-specific sensitivity check showed that the two age groups did *not* deviate significantly from reference in France, Germany, Japan, Poland, and Spain. In the U.S., only older participants showed a significantly lower mean, while the younger ones did not deviate beyond chance-level. Given sample size, the null hypothesis is acceptable because any detectable effect sizes are basically irrelevant for these group comparisons.

The relevance of IC as a construct—and the ICS's sensitivity to group differences—becomes evident when comparing means across educational groups and countries. Using the group with college/vocational training as the reference group, a university degree was not associated with significantly higher IC levels. By contrast, a significantly lower latent mean resulted for participants with lower education level (Δ = −0.26, p < .001). The same pattern featured in Germany and France. In Japan, Spain, and the U.S., it was the highest formal education level that went along with a significantly higher IC level, while the lowest formal education level did not differ significantly from the reference group. In Poland, no group differed significantly from medium-level education. Such differences in trait IC (as predictor variable) are likely to feature at the job market and have economic repercussions for household income (as a criterion), because the numerical differences represent true differences, as the invariance testing ruled out mere measurement bias as an explanation.

The absence of measurement bias is particularly relevant for the cross-national comparison. Though we did not form specific hypotheses for countries, we explored their latent means in reference to the U.S. sample. Latent means were significantly higher in Spain (Δ = 0.26), Poland (Δ = 0.20) and France (Δ = 0.10), yet lower in Germany (Δ = −0.27) and notably in Japan (Δ = −0.99; all ps < .001; for France: p = .04). In the presence of comparable measurement *within* each country and *across* each of the six languages, the discrepancies in the PIAAC pilot samples reveal substantial cross-country differences in IC levels. The differences exceeded those for any other grouping variable we had used for investigating group differences. Given the joint scaling of the latent variable across countries and having set the standard deviation for the reference group to unity, the maximum (Z-scored) distance possible between the countries reflects a *very large* effect size (Cohen's d = 1.25). These results corroborate that the ICS is more than sufficiently sensitive to detect relevant IC discrepancies across countries.

Discussion

We analyzed the psychometric properties of a six-item intellectual curiosity scale in samples taken from six countries and national languages. Our results show that the ICS possesses good psychometric properties, attesting to its essential

---

[9]The RIASEC model of vocational interests (Holland, 1997) allows correlating the ICS with ipsative (individually mean-centered) scale scores. The pattern conformed to the circumplex (hexagon) structure of the RIASEC model. ICS correlated weakly but positively with Investigative (Inquisitive), Artistic, Enterprising (Entrepreneurial), rs = .07–.11, though negatively with Realistic and Conventional Interests, rs = −.15 and −.24.

unidimensionality and factorial validity, while achieving excellent reliability and high levels of measurement invariance across demographic segments and countries, as well as construct validity within each country.

Reliability

First of all, we consider the six-item ICS to be highly reliable, regardless of country ($\omega \approx .90$). Comparing the ICS to the corresponding BFI-2 facet shows that their reliability levels play in completely different leagues ($\omega \approx .60$). Even if an equal number of six rather than four items were used for measuring IC with the BFI-2, a Spearman-Brown corrected estimate would yield $Rel_{corr} \approx .70$). The proximity of country-specific coefficients corroborates that ICS reliability is also akin across languages, whereas reliability of the BFI-2's IC facet meandered ($.46 \leq \omega \leq .71$). In the absence of essential tau-equivalence for the ICS, scale reliability cannot be estimated by Cronbach's α, yet conclusions would be similar: Regarding reliability, the ICS outperforms the BFI-2's IC facet. In the future it seems worthwhile to inspect test-retest reliability: A skill-like trait should display temporal consistency in future assessments, say, across reasonable periods (e.g., less than a year).

Factorial validity and measurement invariance

In terms of factorial validity, the ICS is essentially unidimensional and requires controlling for an association between two items that are partly redundant beyond their association with IC. Improving fit by adding an error covariance IC4–IC5 to the measurement model introduces an exploratory element into CFA. The lack of a theoretical basis for specifying this covariance in advance is compensated by the strictest cross-validation imaginable: replicating the adjusted model with measurement parameters constrained to equality across gender, age, education, and language.

Regarding measurement invariance, MG-CFA supports the generic applicability of the adjusted measurement model. Within countries the ICS is strictly invariant—a rare finding—with the notable exception of scalar non-invariance in Poland (for age groups). Across all countries, metric invariance holds consistently, so respondents share an understanding of ICS content and express their standing in (latent) IC on manifest items using the same psychological units, which allows for comparisons of covariance-based analyses. By contrast, numerical comparisons of mean scores require freeing some item intercepts lest bias be introduced. The ICS is partially scalar invariant with the items IC1, IC2, and IC3 varying in difficulty (Byrne et al., 1989; Cieciuch & Davidov, 2015). Despite such a finding being common (e.g., Dong & Dumas, 2020), partial scalar MI is suitable for exploring psychometric properties (and structural elements such as latent means) in international contexts via properly specified latent variables. And within all countries, due to strict MI for virtually every group, even manifest scale scores are comparable (yet we caution that sum scores are proxies that may be biased if parallel measurement is violated; McNeish & Wolf, 2020).

The replicability of the ICS factor structure across six countries is even more striking if we compare it to traditional scales. The ICS outperformed the BFI-2 IC facet, as the latter merely achieved metric MI. Taking *Typical Intellectual Engagement* as an example, to achieve a fitting measurement model that would allow testing MI, five items had to be removed from the scale (Schroeders et al., 2015). Then, and only then, were the researchers able to attain strict MI across Gender (and School Tracks)—for the modeled latent variable, not for the TIE scale as such. Taking *Need for Cognition* as another example, numerous factor structures have been suggested: For instance, a single factor has been supposed to underlie the 18-item NfC scale (at least for undergraduates), despite an emerging second factor (Sadowski, 1993; see also Davis et al., 1993). Furnham and Thorne (2013) suggested that positively rewording all 18 items would restore unidimensionality. But then, a replicable factor structure that applies to the 18-item NfC scale, as well as to its shortened and extended cousins, is out of reach. Factorial validity (besides clarity of concepts) in the sense of a replicable measurement model is the *conditio sine qua non* for providing the starting point for cross-national invariance.

Construct validity

In terms of convergent validity, IC as measured with the ICS relates to pertinent Big Five personality traits and facets as expected. Generally, the BFI-2 domain *Open-Mindedness* and its IC facet showed the highest correlations, whereas *Negative Emotionality* was least associated. Thus, the ICS associations with the BFI-2 mirror previous relationships (Grüning & Lechner, 2023; Kashdan et al., 2020; Soto & John, 2017). Associations with *Perseverance* and *Sensation Seeking* are compatible with the notion of relevance of IC for academic and job performance, but the ICS is more specific for IC than for these constructs. The ICS is associated with *Job Orientations/Learning Opportunities* substantially and with *Inquisitive Vocational Interests* moderately, both being aspects relevant for career choices. Though not discussed in detail, the ICS is clearly distinguished from theoretically unrelated constructs such as *Traditionalism* (Grüning & Lechner, 2023; Hensley et al., 2012; Soto and John (2017) and *Social Trust*, supporting discriminant validity.

Sensitivity to (known-)group differences

Our findings demonstrate that the ICS is basically gender-fair, and the proximity to a nil effect for gender explains the fluctuating gender differences as they exist in the IC-literature. As for age trends, whereas the pooled sample seemingly confirmed the age trend described in the literature, taking all the empirical evidence together contradicted such a theoretical expectation. It was specifically the U.S. sample that contributed to this trend. Previously reported age trends for IC may have been the outflow of country-specific trends or the result of pooling across heterogeneous samples, resulting in an artifact rather than reflecting a human universal. Alternatively, previous sampling processes may have

unwittingly incurred age-dependent self-selection—a bias that was probably circumvented by a more systematic sampling strategy in the PIAAC context. Also, previously used IC scales may not have been as thoroughly constructed as the ICS, putting the scores of the older population at peril, even in proper random samples.

Looking at IC for three educational levels, the findings are country-specific, so that we could not identify any general trend. Sometimes higher education was associated with increased IC levels compared to medium-level education; at other times, IC levels for basic education groups differed compared to medium-level education. Such inconsistencies are likely due to the fundamental differences in the educational systems of the countries involved, and they indirectly confirm the ICS's sensitivity to pick up such nuanced discrepancies. Compare this to the mean-level differences we found for countries as such: The ICS conveyed large discrepancies across the range of countries analyzed here, supporting its suitability for large-scale assessment and cross-national comparisons.

Limitations

Despite overall convincing findings, the ICS measurement model would profit from future validation in additional, specifically more non-European, samples. Replication would allay concerns for good about the unforeseen residual correlation (IC4–IC5). Researchers and practitioners concerned about the adjusted model may consider dropping either IC4 or IC5 (or paraphrasing one of them) to better approximate strict unidimensionality. Note that this strategy would require new validation efforts (and a shortened five-item instrument might best indicate the number of items in the label to distinguish the ICS-5 from the ICS-6 presented here). Deselecting either item IC4 or IC5 is tricky though. Their average factor loadings are nearly identical. IC4 was inconspicuous in terms of equal intercepts across countries, while equality-constrained IC5 intercepts were prone to misfit. And yet, in terms of linguistic complexity, IC5 appears preferable over IC4. Likewise, the conditional clause in the wording of IC4 may undermine the item's validity for people who subjectively experience a lack of understanding rather infrequently. Whatever item be dropped, any five-item variant is likely to even better approximate unidimensionality than does the current ICS.

The problem of post hoc modification—by freeing intercepts—applies to partial scalar MI too. We did not have any hunch about which items would prove non-invariant, but no obvious a posteriori explanation for these specific items came to mind either. Hence, the question remains: Would the current specifications survive a global study (Bentler & Chou, 1987; Brown & Moore, 2012)?

Another limitation of our MI approach is the repeated testing of nested MI models within groups, plus the repetition of the procedure across multiple grouping variables. Our analyses might have been more informative with intersecting groups, that is, the simultaneous analysis of multiple smaller subgroups (e.g., French males 16–29 years old with a college degree etc.). Such a procedure was impossible given the limited sample sizes and their imbalanced compositions. Note that we are less concerned about p-values and proper Type-I error control here, but about an infinite number of possibly relevant categories that one can compare.

As regards (external) construct validity, it is limited to the variables in the primary data collection. In this regard, our secondary data analysis has both strengths and weaknesses. As for strengths, the associations between IC and *Inquisitive Vocational Interests* or *Job Orientations* border the prediction of relevant (self-reported) job-related outcomes. As for weaknesses, associations with conceptually related, yet more specific constructs in the nomological net are beyond the scope of the available data. Testing the specificity of the ICS (or its incremental validity), and comparing it to related constructs often encountered such as *Need for Cognition* (Cacioppo et al., 1984) or *Typical Intellectual Engagement* (Goff & Ackerman, 1992), potentially also *Epistemic Curiosity* (Litman, 2008), remains a future research avenue. Similarly, Kashdan et al. (2020) advocated six curiosity facets that may be suitable to demarcate IC from other curiosity aspects. The overlap with these existing scales could not be assessed in our study.

A question related to (internal) construct validity concerns the potential advantages of using reversed items. Changing the keying of existing items (or providing additional items with the opposite wording direction) may mitigate bias introduced by acquiescence response-style differences. There can be no indiscriminate recommendation for this practice. Bias control works only if one can inquire about opposite poles while tapping into the same construct (perfect antonyms would be ideal). It is unclear if this is viable for IC as a construct and for the ICS items specifically. An ICS scale that was partially balanced (incomplete acquiescence control) would introduce bias in scores, decreasing the utility for brief measurement in large-scale assessment situations. Ultimately, reverse-keyed items tend to introduce method variance, undermining the goal of unidimensional measurement (Furnham & Thorne, 2013). In the case of the NfC scale, an avoidance factor results that is not fully congruent with the intellectual approach tendency in a two-factor model, alternatively an orthogonal method factor emerges besides a general factor. (We allude here to the "validity of reversed-keyed items" crisis evolving around the assessment of the prominent construct *Growth Mindset*; see Rammstedt et al., 2022; Scherer & Campos, 2022; Lou & Li, 2023; Yeager & Dweck, 2020.) Interested researchers may want to inspect factorial validity and construct validity when introducing inverted ICS items.

Let us address a potential concern about the redundancy between ICS and BFI-2. One vital difference between the ICS and the IC facet in the BFI-2 is that the latter measure cannot be isolated from measuring other personality measures, not without changing the item context. This logic extends to the other facets underlying *Open-Mindedness* as well as completely different personality dimensions. A closer inspection also reveals that the BFI-2 does not represent definitional aspects as fully as the ICS does. The BFI-2 may measure an IC-proxy in the context of other personality traits quite economically, albeit not precisely. The BFI-2 facet approaches the aspect of tackling intricate problems

rather indirectly (i.e., "Avoids intellectual, philosophical discussions"; "Is complex, a deep thinker"), without inquiring about whether one solves the puzzles, or whether one engages in curiosity-related activities and indeed likes them (e.g., "Has little interest in abstract ideas"). Instead, the BFI-2 tends to assess curiosity more generally (e.g., "Is curious about many different things") and may relate to the conceptually broader dimension of being open and interested (in various things). In our view, this partially explains why even the disattenuated correlation coefficient between the two measures is far from perfect (and why this is unlikely to ever be the case)—the scales do not capture the same construct. A comparison with the BFI-2 facet *Intellectual Curiosity* suggests that the strong (though far from perfect) latent correlation with the ICS (.76) does not obviate higher predictive validity of the ICS: Overall, the scale correlations (ICS vs. BFI-2 IC) showed higher convergent validity for the ICS and lower discriminant validity for the BFI-2 IC (double-disattenuated counterparts, corrected for both scales' unreliability [ω], in parentheses): $rs_{(\text{ICS vs BFI-2-IC})}$ = .54 vs. 54 (.65 vs. .79) for *Open-Mindedness:Creative Imagination*; rs = .36 vs. 50 (.42 vs. .71) for *Open-Mindedness:Aesthetic Sensitivity*; rs = .57 vs. .36 (.69 vs. .53) for *Perseverance*; rs =.39 vs. .31 (.46 vs. .44) for *Sensation Seeking*; rs = .53 vs. .33 (.65 vs. .49) for *Learning Opportunities*; rs =.37 vs. .30 (.43 vs. .42) for *Inquisitive Vocational Interests*; and rs = .16 vs. −.02 (.20 vs. −.03) for *Traditionalism*, and rs = −.05 vs. .02 (−.07 vs. .03) for *Social Trust*. All in all, the ICS serves its purpose very well, without being redundant with the BFI-2's IC facet, which in turn seems less specific than the ICS in its pattern of convergent and discriminant correlations.

Let us conclude by pointing out other limitations that can only be addressed by future research. Longitudinal studies could make at least a twofold contribution by inspecting test-retest reliability and stability of the factor structure besides the changes in mean levels across time. Simultaneously, future research should amend the nomological network by including direct competitors and reach beyond self-reports by adding peer-reports or behavioral observations. Such a comprehensive study would allow exploring the nomological net and predictive validity further.

Outlook

Our contribution is but a first, though essential, step toward a comprehensive cultural comparison of IC as the most prominent aspect of curiosity: With IC being an acknowledged central human trait (Maslow, 1943; Peterson & Seligman, 2004), further cross-cultural explorations are needed to firmly establish the generalizability of IC. Hence, future research is to employ the ICS in more diverse cultures, transcending by far the borders of the OECD countries that were available for our secondary data analysis.

While our work is a good starting point toward establishing the ICS as a standard measure of IC that is comparable across countries, it might also help foster and disentangle related concepts and establishing theoretical differences

between them more clearly from a cultural perspective. Given the abundance of research related to intellectual curiosity, constructs such as *Need for Cognition*, *Typical Intellectual Engagement*, and *Epistemic Curiosity* might be subsumed under the broad, inclusive, and holistic umbrella term *Intellectual Curiosity*, with little differentiation between the constructs ("because they are all measuring virtually the same thing," as one reviewer suggested). An alternative view can be delineated to Powell and colleagues' (2016) integrative factor-analytic approach. These authors scrutinized IC factor-analytically across the items of multiple scales and thereby distinguished several dimensions of how intellectual curiosity may be satisfied (e.g., by problem solving or abstract thinking). Moreover, they found that *Epistemic Curiosity* is quite specific in its relationship to IC. While instruments that target IC typically emphasize engagement in complex tasks and enjoying these activities (as NfC and TIE do), *EC* predominantly focuses on the outcome of learning activities (attributed to cognitive deprivation by Powell and colleagues). In contrast, NfC and TIE showed overlap across several dimensions (e.g., intellectual avoidance, problem solving, and abstract thinking). Our preliminary conclusion is: Conceptual weaknesses and overlapping item content require renewed psychometric effort to assess neighboring constructs with higher-than-extant specificity. Continuing the work begun by Woo et al. (2007), Mussel (2010), and Powell et al. (2016), the next step is to locate the ICS in the nomological network with other prominent instruments to explore overlap and uniqueness. As curiosity is widely regarded as a human universal, any theoretical progress in disentangling the related concepts strongly depends on conclusive evidence across cultures and socio-demographics. In this regard, the ICS is setting standards for measuring IC: a reliable, valid, and comparable measurement with six items only. Other scales need to achieve an equal psychometric footing and then transcend single-language findings and cultural specificities before any integrative factor-analytic progress may become tangible and replicable.

Conclusion

Intellectual curiosity is a core facet of the Big Five domain Openness to Experience (or Open-Mindedness in the BFI-2 terminology) and plays a prominent role across many research fields, albeit often under different labels. To advance the measurement of this trait, here we comprehensively assessed the psychometric properties of the six-item Intellectual Curiosity Scale (ICS) in six culturally diverse countries (Japan, Germany, France, Spain, Poland, and the U.S.) using secondary data analysis from the OECD PIAAC pilot studies. Notwithstanding necessary research on the narrower nomological net, our results suggest that the brief and broad 6-item ICS exhibits excellent psychometric properties in terms of unidimensionality, reliability, factorial validity, construct validity as well as comparability (i.e., measurement invariance) across countries. Based on our results, the measure commends itself as especially useful for research purposes in measuring intellectual curiosity,

especially in cross-national settings. We recommend the ICS for research into the precursors, developmental trajectories, and consequences of intellectual curiosity, and likewise for the prediction of relevant outcomes in life.

Authors' contributions

M.B.: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing—original draft, and Writing—review & editing. L.E.: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing—original draft, and Writing—review & editing. D.J.G.: Conceptualization, Writing—original draft, and Writing—review & editing. C.M.L.: Conceptualization, Supervision, and Writing—review & editing.

Declaration of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Part of the research was prepared by L.E. in fulfillment of requirements for a bachelor thesis, supervised by M.B. as a visiting professor (in the capacity of an interim professor) at the Technical University Darmstadt (Germany) in 2021.

ORCID

Matthias Bluemke http://orcid.org/0000-0003-1493-7462
Lukas Engel http://orcid.org/0000-0001-7544-1247
David J. Grüning http://orcid.org/0000-0002-9274-5477
Clemens M. Lechner http://orcid.org/0000-0003-3053-8701

References

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, 11(2), 150–166. https://doi.org/10.1177/1088868306294907

Avery, D. R., Tonidandel, S., Thomas, K. M., Johnson, C. D., & Mack, D. A. (2007). Assessing the multigroup ethnic identity measure for measurement equivalence across racial and ethnic groups. *Educational and Psychological Measurement*, 67(5), 877–888. https://doi.org/10.1177/0013164406299105

Bentler, P. M., & Bonett, D. (1980). Significance tests and goodness-of-fit in analysis of covariance structures. *Psychological Bulletin*, 88(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16(1), 78–117. https://doi.org/10.1177/0049124187016001004

Berlyne, D. E. (1954). A theory of human curiosity. *British Journal of Psychology (London, England: 1953)*, 45(3), 180–191. https://doi.org/10.1111/j.2044-8295.1954.tb01243.x

Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, 44(Suppl 3), S176–S181. https://doi.org/10.1097/01.mlr.0000245143.08679.cc

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361–379). Guilford.

Buchholz, J., & Hartig, J. (2020). Measurement invariance testing in questionnaires: A comparison of three multigroup-CFA and IRT-based approaches. *Psychological Test and Assessment Modeling*, 62(1), 29–53.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, 20(4), 872–882.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed). Routledge.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456

Byrne, B. M., & van de Vijver, F. J. R. (2017). The maximum likelihood alignment approach to testing for approximate measurement invariance: A paradigmatic cross-cultural application. *Psicothema*, 29(4), 539–551. https://doi.org/10.7334/psicothema2017.178

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 119(2), 197–253. https://doi.org/10.1037/0033-2909.119.2.197

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The Efficient assessment of need for cognition. *Journal of Personality Assessment*, 48(3), 306–307. https://doi.org/10.1207/s15327752jpa4803_13

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

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, 95(5), 1005–1018. https://doi.org/10.1037/a0013193

Cheung, G. W., & Lau, R. S. (2012). A direct comparison approach for testing measurement invariance. *Organizational Research Methods*, 15(2), 167–198. https://doi.org/10.1177/1094428111421987

Chu, L., Tsai, J. L., & Fung, H. H. (2021). Association between age and intellectual curiosity: The mediating roles of future time perspective and importance of curiosity. *European Journal of Ageing*, 18(1), 45–53. https://doi.org/10.1007/s10433-020-00567-6

Cieciuch, J., & Davidov, E. (2015). Establishing measurement invariance across online and offline samples. A tutorial with the software packages Amos and Mplus. *Studia Psychologica*, 15(2), 83–99. https://doi.org/10.21697/sp.2015.14.2.06

Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, 13(6), 653–665. https://doi.org/10.1016/0191-8869(92)90236-I

Davidov, E., Datler, G., Schmidt, P., & Schwartz, S. H. (2018). Testing the invariance of values in the Benelux countries with the European Social Survey: Accounting for ordinality. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (2nd ed., pp. 157–179). Routledge.

Davis, T. L., Severy, L. J., Kraus, S. J., & Whitaker, J. M. (1993). Predictors of sentencing decisions: The beliefs, personality variables, and demographic factors of juvenile justice. *Journal of Applied Social Psychology*, 23(6), 451–477. https://doi.org/10.1111/j.1559-1816.1993.tb01098.x

Dellenbach, M., Zimprich, D., & Martin, M. (2008). Kognitiv stimulierende Aktivitäten im mittleren und höheren Erwachsenenalter - ein gerontopsychologischer Beitrag zur Diskussion um informelles Lernen. In A. Kruse (Ed.), *Weiterbildung in der zweiten Lebenshälfte* (pp. 121–159). Bertelsmann.

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

Duncan, T. G., & McKeachie, W. J. (2005). The making of the motivated strategies for learning questionnaire. *Educational Psychologist*, 40(2), 117–128. https://doi.org/10.1207/s15326985ep4002_6

Engelhard, G., & Monsaas, J. A. (1988). Grade level, gender, and school-related curiosity in urban elementary schools. *The Journal of Educational Research*, 82(1), 22–26. https://doi.org/10.1080/00220671.1988.10885860

Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology*, 10, 1507. https://doi.org/10.3389/fpsyg.2019.01507

Fletcher, T. D. (2012). *QuantPsyc: Quantitative psychology tools*. R package version 1.5. https://cran.r-project.org/package=QuantPsyc

Furnham, A., & Thorne, J. D. (2013). Need for cognition. *Journal of Individual Differences*, 34(4), 230–240. https://doi.org/10.1027/1614-0001/a000119

GESIS. (2021). *PIAAC pilot studies on non-cognitive skills*. https://www.gesis.org/piaac/fdz/daten/piaac-pilot-studies-on-non-cognitive-skills

Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology*, 84(4), 537–552. https://doi.org/10.1037/0022-0663.84.4.537

Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every "one" matters. *Psychological Methods*, 6(3), 258–269. https://doi.org/10.1037/1082-989x.6.3.258

Gorges, J., Koch, T., Maehler, D. B., & Offerhaus, J. (2017). Same but different? Measurement invariance of the PIAAC motivation-to-learn scale across key socio-demographic groups. *Large-Scale Assessments in Education*, 5(1), 1–28. https://doi.org/10.1186/s40536-017-0047-5

Grosjean, P., & Ibanez, F. (2018). *Pastecs: Package for analysis of space-time ecological series*. R package version 1.3.21. https://CRAN.R-project.org/package=pastecs

Gross, J., & Ligges, U. (2015). *Nortest: Tests for normality*. R package version 1.0-4. https://CRAN.R-project.org/package=nortest

Grossnickle, E. M. (2016). Disentangling curiosity: Dimensionality, definitions, and distinctions from interest in educational contexts. *Educational Psychology Review*, 28(1), 23–60. https://doi.org/10.1007/s10648-014-9294-y

Grüning, D. J., & Lechner, C. M. (2023). Measuring six facets of curiosity in Germany and the UK: A German-language adaptation of the 5DCR and its comparability with the English-language source version. *Journal of Personality Assessment*, 105(2), 283–295. https://doi.org/10.1080/00223891.2022.2057318

Harackiewicz, J. M., Barron, K. E., Carter, S. M., Lehto, A. T., & Elliot, A. J. (1997). Predictors and consequences of achievement goals in the college classroom: Maintaining interest and making the grade. *Journal of Personality and Social Psychology*, 73(6), 1284–1295. https://doi.org/10.1037/0022-3514.73.6.1284

Hardy, J. H., III, Ness, A. M., & Mecca, J. (2017). Outside the box: Epistemic curiosity as a predictor of creative problem solving and creative performance. *Personality and Individual Differences*, 104, 230–237. https://doi.org/10.1016/j.paid.2016.08.004

Harkness, J. A. (2003). Questionnaire translation. *Cross-Cultural Survey Methods*, 1, 35–56.

Hensley, B., Martin, P., Margrett, J. A., MacDonald, M., Siegler, I. C., Poon, L. W., Jazwinski, S. M., Green, R. C., Gearing, M., Woodard, J. L., Johnson, M. A., Tenover, J. S., Rodgers, W. L., Hausman, D. B., Rott, C., Davey, A., & Arnold, J., The Georgia Centenarian Study 1. (2012). Life events and personality predicting loneliness among centenarians: Findings from the Georgia centenarian study. *The Journal of Psychology*, 146(1–2), 173–188. https://doi.org/10.1080/00223980.2011.613874

Henze, N., & Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods*, 19(10), 3595–3617. https://doi.org/10.1080/03610929008830400

Holland, J. L. (1997). *Making vocational choices* (3rd ed.). Psychological Assessment Resources.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. https://doi.org/10.1007/BF02289447

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

Johnson, J. A. (2014). Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, 51, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity Activates Reward Circuitry and Enhances Memory. *Psychological Science*, 20(8), 963–973. https://doi.org/10.1111/j.1467-9280.2009.02402.x

Kashdan, T. B., Disabato, D. J., Goodman, F. R., & McKnight, P. E. (2020). The Five-Dimensional Curiosity Scale Revised (5DCR): Briefer subscales while separating overt and covert social curiosity. *Personality and Individual Differences*, 157, 109836. https://doi.org/10.1016/j.paid.2020.109836

Kashdan, T. B., Rose, P., & Fincham, F. D. (2004). Curiosity and exploration: Facilitating positive subjective experiences and personal growth opportunities. *Journal of Personality Assessment*, 82(3), 291–305. https://doi.org/10.1207/s15327752jpa8203_05

Kashdan, T. B., Stiksma, M. C., Disabato, D. J., McKnight, P. E., Bekier, J., Kaji, J., & Lazarus, R. (2018). The Five-Dimensional Curiosity Scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. *Journal of Research in Personality*, 73, 130–149. https://doi.org/10.1016/j.jrp.2017.11.011

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate Normality. *The R Journal*, 6(2), 151–162. https://doi.org/10.32614/RJ-2014-031

Lauriola, M., Litman, J. A., Mussel, P., De Santis, R., Crowson, H. M., & Hoffman, R. R. (2015). Epistemic curiosity and self-regulation. *Personality and Individual Differences*, 83, 202–207. https://doi.org/10.1016/j.paid.2015.04.017

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

Lenth, R. V. (2021). *Emmeans: Estimated marginal means, aka least-squares means*. R package version 1.6.3. https://CRAN.R-project.org/package=emmeans

Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences*, 44(7), 1585–1595. https://doi.org/10.1016/j.paid.2008.01.014

Litman, J. A. (2010). Relationships between measures of I- and D-type curiosity, ambiguity tolerance, and need for closure: An initial test of the wanting-liking model of information-seeking. *Personality and Individual Differences*, 48(4), 397–402. https://doi.org/10.1016/j.paid.2009.11.005

Litman, J. A., & Jimerson, T. L. (2004). The measurement of curiosity as a feeling of deprivation. *Journal of Personality Assessment*, 82(2), 147–157. https://doi.org/10.1207/s15327752jpa8202_3

Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment*, 80(1), 75–86. https://doi.org/10.1207/S15327752JPA8001_16

Lou, N. M., & Li, L. M. W. (2023). The mindsets × societal norm effect across 78 cultures: Growth mindsets are linked to performance weakly and well-being negatively in societies with fixed-mindset norms. *British Journal of Educational Psychology*, 93(1), 134–152. https://doi.org/10.1111/bjep.12544

Maehler, D. B., & Rammstedt, B. (Eds.). (2020). *Large-scale cognitive assessment: Analyzing PIAAC data*. Springer Nature. https://doi.org/10.1007/978-3-030-47515-4

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519–530. https://doi.org/10.1093/biomet/57.3.519

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2018). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, 23(3), 524–545. https://doi.org/10.1037/met0000113

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370–396. https://doi.org/10.1037/h0054346

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology*, 52(6), 1258–1265. https://doi.org/10.1037/0022-3514.52.6.1258

McNeish, D., An, J., & Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personality Assessment*, 100(1), 43–52. https://doi.org/10.1080/00223891.2017.1281286

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, 52(6), 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111–130. https://doi.org/10.21500/20112084.857

Mussel, P. (2010). Epistemic curiosity and related constructs: Lacking evidence of discriminant validity. *Personality and Individual Differences*, 49(5), 506–510. https://doi.org/10.1016/j.paid.2010.05.014

Mussel, P., Spengler, M., Litman, J. A., & Schuler, H. (2012). Development and validation of the German Work-Related Curiosity Scale. *European Journal of Psychological Assessment*, 28(2), 109–117. https://doi.org/10.1027/1015-5759/a000098

Noar, S. M. (2003). The role of structural equation modeling in scale development. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(4), 622–647. https://doi.org/10.1207/S15328007SEM1004_8

Organisation for Economic Co-operation and Development (OECD). (2013). *PISA 2012 results: Ready to learn (Volume III): Students' engagement, drive and self-beliefs*. OECD Publishing. https://doi.org/10.1787/9789264201170-en

Organisation for Economic Co-operation and Development (OECD). (2018a). *Programme for the International Assessment of Adult Competencies (PIAAC), English pilot study on non-cognitive skills*. GESIS Data Archive, Cologne. ZA6940 Data file Version 1.0.0. https://doi.org/10.4232/1.13062

Organisation for Economic Co-operation and Development (OECD). (2018b). *Programme for the International Assessment of Adult Competencies (PIAAC), International pilot study on non-cognitive skills*. GESIS Data Archive, Cologne. ZA6941 Data file Version 1.0.0. https://doi.org/10.4232/1.13063

Organisation for Economic Co-operation and Development (OECD). (2019). What the Survey of Adult Skills (PIAAC) measures. In *The survey of adult skills: Reader's companion* (3rd ed.). OECD Publishing. https://doi.org/10.1787/0e96cba5-en

Orcutt, J. M., & Dringus, L. P. (2017). Beyond being there: Practices that establish presence, engage students and influence intellectual curiosity in a structured online learning environment. *Online Learning*, 21(3), 15–35. https://doi.org/10.24059/olj.v21i3.1231

Partsch, M. V., & Danner, D. (2021). Measuring self-control in international large-scale surveys: Development and validation of a four-item scale in English, French, German, Japanese, Polish, and Spanish. *European Journal of Psychological Assessment*, 37(5), 409–418. https://doi.org/10.1027/1015-5759/a000618

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. Oxford University Press.

Piotrowski, J. T., Litman, J. A., & Valkenburg, P. (2014). Measuring epistemic curiosity in young children. *Infant and Child Development*, 23(5), 542–553. https://doi.org/10.1002/icd.1847

Powell, C., & Nettelbeck, T. (2014). Intellectual curiosity may not incrementally predict academic success. *Personality and Individual Differences*, 64, 7–11. https://doi.org/10.1016/j.paid.2014.01.045

Powell, C., Nettelbeck, T., & Burns, N. R. (2016). Deconstructing intellectual curiosity. *Personality and Individual Differences*, 95, 147–151. https://doi.org/10.1016/j.paid.2016.02.037

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rammstedt, B. (2013). *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich: Ergebnisse von PIAAC 2012*. Waxmann. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-360687

Rammstedt, B., Grüning, D. J., & Lechner, C. M. (2022). Measuring growth mindset: Validation of a three-item and a single-item scale in adolescents and adults. *European Journal of Psychological Assessment*. Advance online publication. https://doi.org/10.1027/1015-5759/a000735

Raykov, T., & Bluemke, M. (2021). Examining multidimensional measuring instruments for proximity to unidimensional structure using latent variable modeling. *Educational and Psychological Measurement*, 81(2), 319–339. https://doi.org/10.1177/0013164420940764

Revelle, W. (2020). *psych: Procedures for personality and psychological research*. Version = 2.0.12. https://CRAN.R-project.org/package=psych

Rosseel, 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

Royston, J. P. (1983). A simple method for evaluating the Shapiro–Francia W-test of non-normality. *The Statistician)*, 32(3), 297–300. https://doi.org/10.2307/2987935

Royston, P. (1995). Remark AS R94: A remark on algorithm AS 181: The W-test for normality. *Applied Statistics)*, 44(4), 547–551. https://doi.org/10.2307/2986146

Sadowski, C. J. (1993). An examination of the short need for cognition scale. *The Journal of Psychology*, 127(4), 451–454. https://doi.org/10.1080/00223980.1993.9915581

Scherer, R., & Campos, D. G. (2022). Measuring those who have their minds set: An item-level meta-analysis of the implicit theories of intelligence scale in education. *Educational Research Review*, 37, 100479. https://doi.org/10.1016/j.edurev.2022.100479

Schroeders, U., & Gnambs, T. (2018). Degrees of freedom in multigroup confirmatory factor analyses. *European Journal of Psychological Assessment*, 36(1), 105–113. https://doi.org/10.1027/1015-5759/a000500

Schroeders, U., Schipolowski, S., & Böhme, K. (2015). Typical intellectual engagement and achievement in math and the sciences in secondary education. *Learning and Individual Differences*, 43, 31–38. https://doi.org/10.1016/j.lindif.2015.08.030

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. https://doi.org/10.2307/2333709

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). *Afex: Analysis of factorial experiments*. R package version 1.0-1. https://CRAN.R-project.org/package=afex

Smith, M. C., Rose, A. D., Ross-Gordon, J., & Smith, T. J. (2015). *Adults' readiness to learn as a predictor of literacy skills*. Retrieved September 02, 2022, from https://www.researchgate.net/profile/Amy-Rose-11/publication/285588147_Adults'_Readiness_to_Learn_as_a_Predictor_of_Literacy_Skills/links/5660696308ae418a78665846/Adults-Readiness-to-Learn-as-a-Predictor-of-Literacy-Skills.pdf

Soto, C. J., & John, O. P. (2009). Ten facet scales for the Big Five Inventory: Convergence with NEO PI-R facets, self-peer agreement, and discriminant validity. *Journal of Research in Personality*, 43(1), 84–90. https://doi.org/10.1016/j.jrp.2008.10.002

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117–143. https://doi.org/10.1037/pspp0000096

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78–107. https://doi.org/10.1086/209528

Steiner, M. D., & Grieder, S. (2020). EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software*, 5(53), 2521. https://doi.org/10.21105/joss.02521

Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using M plus and the lavaan/semTools packages. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1), 111–130. https://doi.org/10.1080/10705511.2019.1602776

Trudewind, C. (2000). Curiosity and anxiety as motivational determinants of cognitive development. In J. Heckhausen (Ed.), *Advances in psychology* (Vol. 131, pp. 15–38). https://doi.org/10.1016/S0166-4115(00)80004-7

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70. https://doi.org/10.1177/109442810031002

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41–71). Springer US. https://doi.org/10.1007/978-1-4615-4397-8_3

von Stumm, S., & Ackerman, P. L. (2013). Investment and intellect: A review and meta-analysis. *Psychological Bulletin*, 139(4), 841–869. https://doi.org/10.1037/a0030746

von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science*, 6(6), 574–588. https://doi.org/10.1177/1745691611421204

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, *54*(3), 426–482. https://doi.org/10.1090/S0002-9947-1943-0012401-3

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*(4), 669–689. https://doi.org/10.1016/S0191-8869(00)00064-7

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data manipulation*. R package version 1.0.5. https://CRAN.R-project.org/package=dplyr

Wiggin, K. L., Reimann, M., & Jain, S. P. (2019). Curiosity tempts indulgence. *Journal of Consumer Research*, *45*(6), 1194–1212. https://doi.org/10.1093/jcr/ucy055

Woo, S. E., Harms, P. D., & Kuncel, N. R. (2007). Integrating personality and intelligence: Typical intellectual engagement and need for cognition. *Personality and Individual Differences*, *43*(6), 1635–1639. https://doi.org/10.1016/j.paid.2007.04.022

Yeager, D. S., & Dweck, C. S. (2020). What can be learned from growth mindset controversies? *The American Psychologist*, *75*(9), 1269–1284. https://doi.org/10.1037/amp0000794

Zagumny, M. J. (2016). Q-test of undergraduate epistemology and scientific thought: Development and testing of an assessment of scientific-epistemology. *International Journal of Psychological and Behavioral Sciences*, *10*(5) 1563–1569. https://doi.org/10.5281/zenodo.1124261

Zagumny, M. J. (2018). Psychometric examination of the QUEST-25: An online assessment of intellectual curiosity and scientific epistemology. *International Journal of Educational and Pedagogical Sciences*, *12*(7), 928–932. https://doi.org/10.5281/zenodo.1317376

Zimprich, D., Allemand, M., & Dellenbach, M. (2009). Openness to experience, fluid intelligence, and crystallized intelligence in middle-aged and old adults. *Journal of Research in Personality*, *43*(3), 444–454. https://doi.org/10.1016/j.jrp.2009.01.018