METHODOLOGICAL CHALLENGES IN THE ASSESSMENT OF GOOD CHARACTER

vom Promotionsausschuss des Fachbereichs Psychologie der Rheinland-Pfälzischen Technischen Universität Kaiserslautern-Landau zur Verleihung des akademischen Grades Doktor der Philosophie (Dr. phil.) genehmigte

DISSERTATION

von

Melanie Viola Partsch, M.Sc.

Vorsitzende des Promotionsausschusses und der Promotionskommission:
Prof. Dr. Tanja Lischetzke

Berichterstatter:
Dr. Clemens M. Lechner
Prof. Dr. Manfred Schmitt

Datum der wissenschaftlichen Aussprache:
10. November 2023

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

While working on this dissertation, I was privileged to experience the support of many people. I would like to thank you all wholeheartedly.

Foremost, I would like to express my sincere gratitude to my supervisor and mentor, Clemens Lechner. You have been a source of great encouragement, support, infectious enthusiasm, and inspiration. Working with you has been a pleasure, and I am deeply grateful for everything I have learned from you during the past years.

Furthermore, I would like to sincerely thank my second supervisor, Manfred Schmitt, for encouraging and inspiring my research in the form of valuable discussions and comments and for always promptly responding to my questions and queries.

I would also like to warmly thank my colleagues from the past years—foremost the teams *Scale Development and Documentation* and *Questionnaire Design and Evaluation* at GESIS, Matthias Blümke, Gabriel Olaru, and Gerard Saucier—for the supportive atmosphere and many stimulating discussions.

Last but not least, I would like to express my heartfelt gratitude to my family and friends. I am deeply thankful to my whole family—especially my parents Heidi and Rainer, my brother Fabian, and my grandmother Traude—for their unconditional love, support, and belief in me. Furthermore, I am more than grateful to my friends that I was privileged to have met on my life's journey so far for all the support, encouragement, love, precious conversations, and refreshing distraction. Laura and Hedda, always knowing you by my side during the past years has been invaluable.

SUMMARY

Since the turn of the millennium, character research has been on the rise among psychological researchers. In 2004, the field of positive psychology introduced the Values in Action (VIA) framework encompassing 24 theoretically justified and empirically supported character strengths intended for the measurement of good character. Their assignment to six "core virtues" according to Linnaean principles links the 24 character strengths to philosophical and religious theories of virtue. However, the originally developed proprietary VIA Inventory of Strengths (VIA-IS) for the measurement of the 24 character strengths and its public domain counterpart, the IPIP-VIA, are based on a relatively crude scale development approach. Yet, the VIA-IS and the IPIP-VIA dominated (applied) character research for a long time. While researchers recently refined the proprietary VIA instruments, no character strength scales developed according to the state of the art are available in the public domain, thwarting progress in character research. Furthermore, most factor-analytic studies on the hierarchical structure of the 24 VIA character strengths yielded inconsistent results regarding the number and nature of global VIA constructs due to differing methodological standards and strategies. Only recently, a growing body of research consonantly has suggested that three global constructs span the VIA trait space. Consequently, there is only one proprietary inventory for measuring global VIA constructs and none that is available in the public domain. Against this backdrop, this dissertation addressed three methodological challenges in character assessment, taking an open-science approach, a (cross-country) replicability approach, and an integrative approach (i.e., integrating the results into the larger picture of personality science, particularly linking the VIA character traits to the Big Five and value traits).

Study 1 revised the English-language IPIP-VIA and concurrently translated/adapted it to German to yield character strength scales especially suitable for cross-cultural large-scale assessment: The 96-item IPIP-VIA-R measures each character strength with four balanced-keyed, content-valid, and cross-culturally adaptable items building scales that showed satisfactory reliability, (partial) scalar measurement invariance across Germany and the UK, and evidence of construct and criterion validity. Study 2 applied the IPIP-VIA-R and a rigorous factor-analytic approach to revisit the hierarchical structure of the 24 VIA character strengths, revealing three well-interpretable global “core strengths” that were replicable across Germany and the UK: positivity, dependability, and mastery. Study 3 applied an Ant Colony Optimization algorithm to select an optimal 18-item subset of the IPIP-VIA-R to measure each core strength with a balanced-keyed, content-valid six-item scale that again showed satisfactory reliability, scalar measurement invariance across Germany and the UK, and evidence of construct and criterion validity.

Taken as a whole, the dissertation advanced the measurement of VIA character traits in the public domain, the understanding of the VIA character trait space (especially its intersection with Big Five personality and basic human values), and the establishment of the VIA trait hierarchy. To address its research questions framed as methodological challenges, the dissertation introduced and elaborated methodological approaches that researchers might adapt to other individual differences constructs. Even though there remain challenges to be taken up in future work (e.g., adapting the IPIP-VIA-R character and core strength scales for use in a more diverse set of cultures; multi-informant assessment), researchers and survey programs can readily apply the character scales developed as part of this dissertation.

ZUSAMMENFASSUNG

Seit der Jahrtausendwende ist die Charakterforschung in der psychologischen Forschung im Aufstieg begriffen. Im Jahr 2004 wurde in der Positiven Psychologie die Values-in-Action- (VIA-) Klassifikation eingeführt, die 24 theoretisch begründete und empirisch gestützte Charakterstärken zur Messung guten Charakters umfasst. Ihre Zuordnung zu sechs "Kerntugenden" nach dem Linné'schen Prinzip verbindet die 24 Charakterstärken mit philosophischen und religiösen Tugendtheorien. Das ursprünglich entwickelte proprietäre VIA Inventory of Strengths (VIA-IS) zur Messung der 24 Charakterstärken sowie sein frei zugängliches Pendant, das IPIP-VIA, beruhen jedoch auf einem nicht ausgereiften Skalenentwicklungsansatz. Dennoch haben das VIA-IS und das IPIP-VIA die (angewandte) Charakterforschung lange Zeit dominiert. Während Forschende die proprietären VIA-Instrumente in jüngster Zeit weiterentwickelt haben, sind keine nach dem aktuellen Stand der Wissenschaft entwickelten Skalen für Charakterstärken frei zugänglich, was den Fortschritt der Charakterforschung hemmt. Darüber hinaus ergaben die meisten faktorenanalytischen Studien zur hierarchischen Struktur der 24 VIA-Charakterstärken aufgrund unterschiedlicher methodischer Standards und Strategien uneinheitliche Ergebnisse hinsichtlich der Anzahl und Beschaffenheit globaler VIA-Konstrukte. Erst seit Kurzem deuten Forschungsergebnisse vermehrt darauf hin, dass drei globale Konstrukte den VIA-Eigenschaftsraum aufspannen. Folglich gibt es nur ein proprietäres Inventar zur Messung globaler VIA-Konstrukte und keines, das frei zugänglich ist. Vor diesem Hintergrund stellte sich diese Dissertation drei methodischen Herausforderungen in der Charaktermessung und verfolgte dabei einen Open-Science-Ansatz, einen (länderübergreifenden) Replizierbarkeitsansatz sowie einen integrativen Ansatz (Einbindung der Ergebnisse in die Persönlichkeitsforschung: Wie verhalten sich VIA-Charaktereigenschaften zu den Big Five und Werteeigenschaften?).

In Studie 1 wurde das IPIP-VIA überarbeitet und gleichzeitig aus dem Englischen ins Deutsche übersetzt/adaptiert, um Skalen zur Messung von Charakterstärken zu entwickeln,

die sich besonders auch für den Einsatz in großen internationalen Umfrageprogrammen eignen: Das 96 Items umfassende IPIP-VIA-R misst jede Charakterstärke mit vier balancierten, inhaltsvaliden und kulturübergreifend anpassbaren Items. Die resultierenden Skalen wiesen zufriedenstellende Reliabilität, (partiell) skalare Messinvarianz in Deutschland und Großbritannien sowie Evidenz für Konstrukt- und Kriteriumsvalidität auf. In Studie 2 wurden das IPIP-VIA-R sowie ein stringenter faktorenanalytischer Ansatz angewandt, um die hierarchische Struktur der 24 VIA-Charakterstärken zu überprüfen. Dabei wurden drei gut interpretierbare globale "Kernstärken" ermittelt, die über Deutschland und Großbritannien hinweg replizierbar waren: Positivität (*positivity*), Zuverlässigkeit/Beständigkeit (*dependability*) und Meisterung/Souveränität (*mastery*). In Studie 3 wurde ein Ant-Colony-Optimization-Algorithmus angewandt, um ein optimales 18-Item-Subset des IPIP-VIA-R auszuwählen, anhand dessen jede Kernstärke mit einer balancierten, inhaltsvaliden 6-Item-Skala gemessen werden kann. Diese Skalen zeigten zufriedenstellende Reliabilität, skalare Messinvarianz in Deutschland und Großbritannien sowie Evidenz für Konstrukt- und Kriteriumsvalidität.

Insgesamt hat die Dissertation die frei zugängliche Messung von VIA-Charaktereigenschaften, das Verständnis des VIA-Charaktereigenschaften-Spektrums (insbesondere seine Überschneidung mit Big Five-Persönlichkeitseigenschaften und menschlichen Grundwerten) sowie die Etablierung der VIA-Eigenschaftshierarchie vorangetrieben. Um die als methodische Herausforderungen formulierten Forschungsfragen zu beantworten, hat die Dissertation methodische Ansätze eingeführt und weiterentwickelt, die Forschende auf andere differenzialpsychologische Konstrukte übertragen können. Auch wenn in der allgemeinen Diskussion wichtige weitere Herausforderungen aufgezeigt werden (z. B. die Adaption der IPIP-VIA-R Charakter- und Kernstärkenskalen für den Einsatz in vielfältigeren Kulturen; Multi-Informanten-Erhebung), können Forschende und Erhebungsprogramme die in dieser Dissertation entwickelten Skalen unmittelbar anwenden.

GENERAL INTRODUCTION

Virtuousness or “good character” have always occupied mankind. Since ancient times, religious and philosophical intellectuals and thought leaders from all over the world have described and classified virtuous conduct. Cardinal virtues like wisdom or justice have been defined by Plato in ancient Greece and by Confucius in ancient China and have influenced the doctrines of other well-known philosophers from these times like Aristotle, Seneca, and Mencius or later the theological virtues from Christian doctrine.

To date, the topic has not lost its relevance. In addition to contemporary reconstructions of philosophical virtue theories (e.g., Adams, 2006), modern-day psychological research has made both theoretical and empirical contributions to the understanding of human character (e.g., Fowers et al., 2021; McGrath et al., 2018; Peterson & Seligman, 2004; Wright et al., 2020). Psychologists’ interest in human character has grown with the growing recognition and evidence that individual differences in human character are integral to the understanding of human behavior and co-existence: Character traits can contribute to personal growth, well-being, and flourishing, as well as to beneficial social interaction guided by moral principles (e.g., Fowers et al., 2021; Peterson & Seligman, 2004). In a nutshell, character traits have proven to be important guideposts to the “good life”.

Psychological researchers have conceptualized character as measurable individual difference constructs representing desirable, positively—usually also morally—valued traits that can be acquired or developed through personal and institutional practice or training and that manifest in well-motivated, intentional, self- and other-benefitting behavior. Examples of character traits are perspective, integrity, kindness, fairness, forgiveness, and gratitude (e.g., Fowers et al., 2021; Peterson & Seligman, 2004). As positively valued traits, character traits can be localized at the intersection of (basic) personality traits and human values and, as such, represent (personally and socially) growth-oriented personality descriptors. As personal and social resources, character traits have been of great interest in various applied (research) areas, such as clinical psychology (e.g., Rashid, 2015), educational psychology (e.g., Wagner & Ruch, 2015; Weber et al., 2016), or organizational psychology (e.g., Harzer & Ruch, 2014; Lavy & Littman-Ovadia, 2017). Furthermore, as "21st century skills" or "social and emotional skills" constructs similar to character traits are promoted by the National Research Council (2012) or the Organisation for Economic Co-operation and Development (Chernyshenko et al., 2018), underscoring their modern-day societal and political relevance.

Since the turn of the millennium, psychological research on human character has considerably increased due to the introduction of the "Values in Action" (VIA) classification by Peterson and Seligman (2004). VIA allows to measure and describe "good character" based on an extensive collection of 24 *character strengths* that were mapped onto six universal *core virtues* on theoretical grounds. Notably, while enjoying great popularity in the field of positive psychology, the VIA framework has not yet become established in mainstream personality science, which remains wedded to the Big Five and the Big Six. Yet, VIA represents a valuable complement to "basic" personality frameworks: While the Big Five and the Big Six originate in psycho-lexical approaches and describe the individual mainly with (supposedly) neutral traits from different levels of a factor-analytic trait hierarchy, VIA describes "the contributor to the social good" (McGrath et al., 2020, p. 132) based on positively valued traits from a theoretically developed classification that ties in with philosophical and religious traditions on character. However, VIA has not yet met certain methodological standards of personality science. Methodological challenges—especially regarding the measurement of character traits—have only recently gained appropriate attention (e.g., in Feraco et al., 2022; McGrath, 2019), with advancements being limited to proprietary instruments. This is why the present dissertation seeks to meet methodological challenges in the measurement of VIA character with a focus on open science, cross-cultural adaptability and replicability, and integration into the bigger picture of personality science.

Theoretical Background

The VIA Character Framework

The Linnaean Classification of Character Strengths and Virtues. To tie in with established virtue theories, Peterson and Seligman (2004) reviewed various philosophical and religious writings from China (Confucianism and Taoism), South Asia (Buddhism and Hinduism), and the West (ancient Greece, Judeo-Christianity, and Islam) and abstracted *wisdom and knowledge, courage, humanity, justice, temperance,* and *transcendence* as shared universal core virtues. For the measurement of character, they compiled 24 theoretically justified and empirically supported character strengths that they defined as (universally) positively valued, morally relevant, skill- and trait-like, acquired, and cultivated individual difference constructs. These 24 character strengths resulted from a comprehensive list of positively valued characteristics drawn from various contexts (e.g., character development programs, boy scouts, song lyrics) that was systematically winnowed and condensed based on 10 definitional criteria of character constructs (see Study 1). For classification purposes, Peterson and Seligman (2004) mapped the 24 character strengths onto the six core virtues in a *Linnaean* manner—that is, based on common attributes, as Carl Linnaeus proceeded in his biological classification of species. The 24 character strengths are composed of:

- Four *emotional* strengths mapped onto the core virtue of *courage*, in particular industriousness/perseverance, integrity, valor/bravery, and zest
- Three *civic* strengths mapped onto the core virtue of *justice*, in particular citizenship/teamwork, equity, and leadership
- Three *interpersonal* strengths mapped onto the core virtue of *humanity*, in particular capacity for love, kindness, and social intelligence
- Four *protective/regulatory* strengths mapped onto the core virtue of *temperance*, in particular forgiveness, modesty/humility, prudence, and self-regulation

- Five *meaning-providing* strengths mapped onto the core virtue of *transcendence*, in particular appreciation of beauty, gratitude, hope, humor, and spirituality/religiousness
- Five *cognitive* strengths mapped onto the core virtue of *wisdom and knowledge*, in particular curiosity, judgment, love of learning, originality, and perspective

Table 1.1 of Study 1 provides short definitions of each strength. For a more detailed description of the six core virtues and the Linnaean strengths-virtues classification, see online supplement of Study 2 (https://osf.io/vs54k). Based on the Linnaean classification, Peterson and Seligman (2004, p. 13) speculated what constitutes a person of good character: "We speculate that all these [six core] virtues must be present at above-threshold values for an individual to be deemed of good character. […] We are comfortable saying that someone is of good character if he or she displays but 1 or 2 strengths within a virtue group."

On the Moral Relevance of VIA Character Strengths. The majority of character researchers considers the moral relevance of a character trait to be one of its key defining features (Fowers et al., 2021). Consistent with this, Peterson and Seligman (2004) assumed that 22 out of 24 VIA character strengths would be morally valued—only to zest and industriousness/perseverance this would not or only partly apply. Later research additionally identified appreciation of beauty, humor, love of learning, and curiosity as not inherently morally valued character strengths and suggested that they might instead be conditionally morally valued given that they come along with inherently morally valued character strengths (e.g., humor paired with, for example, forgiveness and hope results in morally valued benevolent humor; Peterson & Park, 2009; Ruch & Heintz, 2016; Ruch & Stahlmann, 2020). Probably also because of the missing "moral core" in some VIA character strengths, Fowers et al. (2021) criticized that positive psychology was deprioritizing the concept of morality in their approaches to character.

The Factor-Analytic VIA Trait Hierarchy. Despite being instructive in describing the VIA character trait space, Peterson and Seligman's (2004) *theoretical* Linnaean classification of (more specific) character strengths to abstract core virtues does not conform to the state-of-the-art procedure in individual differences research of establishing a trait hierarchy with different abstraction levels. To establish trait hierarchies for other individual differences constructs, such as intelligence or Big Five and Big Six personality, researchers frequently used *empirical* factor-analytic methods (e.g., Goldberg, 2006; Lang et al., 2016; Lee & Ashton, 2018). In personality trait hierarchies, *facets* are of comparable specificity to the 24 VIA character strengths (McGrath et al., 2020). The (co-)variance in the personality facets can be factor-analytically aggregated to more and more global (i.e., abstract) *aspects*, *domains*, and *metatraits* (e.g., DeYoung, 2006; DeYoung et al., 2007; McCrae & John, 1992).

Since the introduction of the VIA framework, also several factor-analytic studies based on the 24 character strengths have been conducted (e.g., Najderska & Cieciuch, 2018; Noronha et al., 2015; Ruch et al., 2010). However, as detailed in Study 2, these studies yielded inconsistent and often non-replicable results suggesting 1–6 useful more global dimensions because of the great variability in their methodological approaches, quality, and rigor. Only relatively recently, three global dimensions labeled *Caring*, *Inquisitiveness*, and *Self-Control* were replicated across several studies using different samples and VIA instruments (e.g., McGrath, 2015; McGrath et al., 2018, 2022; McGrath & Wallace, 2021)—with a somewhat lax understanding of replicability though (i.e., several of the highly loading character strengths differed across studies). Establishing the VIA trait hierarchy is important for a better understanding of the makeup of (VIA) character—especially relative to Big Five and Big Six personality—and allows measuring character on different levels of specificity versus globality.

Table 1

Comparison of the VIA, the Big Five, and the Big Six framework

	VIA character	**Big Five personality**	**Big Six personality**
Purpose of framework (McGrath et al., 2020; p. 132)	Describe "individual as a contributor to the social good"	"Global description of the individual"	"Global description of the individual"
Developmental approach to framework	Theoretical classification compatible with religious and philosophical virtue theories	Psycho-lexical approach based on sedimentation hypothesis; factor-analytic trait hierarchy	Psycho-lexical approach based on sedimentation hypothesis; factor-analytic trait hierarchy
Input/Content on which constructs are based	Comprehensive collection of positive characteristics	Comprehensive collection of neutral terms (plus, inadvertently, some morality-relevant terms)	Comprehensive collection of prototypically/frequently used terms
Proportion of morality-relevant, highly evaluative input/content	High	Low	Intermediate
Conception of human beings (to be) captured by constructs	Sociopsychic/psychocultural	Neuropsychic	Neuropsychic and some sociopsychic/psychocultural aspects
Correlation-based redundancy [a]			
McGrath et al. (2020)		**1** out of 24 VIA strengths likely redundant with a NEO facet	**8** out of 24 VIA strengths likely redundant with a HEXACO facet
Ruch et al. (2021; Table S5)		**3** out of 24 VIA strengths likely redundant with a NEO facet	

Note. [a] Based on McGrath et al.'s (2020) cutoff of $r \geq .60$ between two constructs for them to be likely redundant.

Contrasting Character With Personality

While the VIA character framework was only introduced in 2004, the starting point for present-day taxonomic personality research was marked by Allport and Odbert's (1936) first systematic large-scale lexical approach (John et al., 1988). Therefore, it is important to contrast VIA character traits with (basic) personality traits of five- or six-factor models (e.g., the BFI or HEXACO domains or facets; Ashton & Lee, 2008; Soto & John, 2017a): I will first compare the VIA character framework with taxonomic personality frameworks regarding their overall purpose and general developmental approach. Second, I will zoom in to compare the constructs of VIA, the Big Five, and the Big Six regarding the content/input they are based on and the conception of human beings they convey. Third, I will summarize empirical results on the redundancy versus distinctness between VIA character and Big Five/Big Six personality constructs. Table 1 provides an overview of these comparisons.

An Empirical Approach to Global Personality Versus a Theoretical Approach to Social Personality (i.e., Character). The purpose of both the Big Five and the Big Six personality framework is to allow for a "global description of the individual" (McGrath et al., 2020, p. 132). To that end, the Big Five and Big Six personality factors were originally identified based on an empirical psycho-lexical approach following the sedimentation hypothesis that assumes that the more salient and socially relevant an individual difference is for a people, the more likely it is recovered as a single word (or adjective) in its language (John et al., 1988; Klages, 1932). Based on this assumption, the first step of the lexical approach is to extract *all* terms that describe personality or individual differences from a dictionary (e.g., Allport and Odbert, 1936, extracted 17,953 terms from *Webster's New International Dictionary of the English Language*, 1928), to winnow the list based on one's demarcation of personality (e.g., Allport and Odbert, 1936, reduced their list to 4,504 "neutral terms designating possible personal traits", p. 38) in a second step, to collect data (i.e., have people describe themselves or others based on these terms) in a third step, and, finally, to apply factor-analytic methods to the data to yield the relevant global dimensions of personality. This last step resulted in numerous studies in either five factors (i.e., the Big Five; e.g., Goldberg, 1990) or six factors (i.e., the Big Six; e.g., Lee & Ashton, 2008; Saucier, 2009). The psycho-lexical Big Five and Big Six, respectively, were expanded both theoretically and empirically based on questionnaire approaches (Strus & Cieciuch, 2019). Questionnaire-based models or inventories are (a) the Five-Factor Model (FFM; e.g., McCrae & Costa, 2008) with the corresponding NEO inventories (Costa & McCrae, 2014), (b) the BFIs (i.e., Big Five Inventories; e.g., John et al., 2008; Soto & John, 2017a), and (c) the HEXACO six-factor model with the corresponding HEXACO inventories (e.g., Ashton & Lee, 2008) and the shorter Questionnaire Big Six scales (QB6; Thalmayer & Saucier, 2014). Part of the expansion of questionnaire-based approaches was to derive more specific facets underlying the global five or six factors in the Big Five or Big Six trait hierarchies (e.g., Lee & Ashton, 2006; Soto & John, 2017a).

By contrast, the purpose of the VIA character framework is to describe the individual as “a contributor to the social good” (McGrath et al., 2020, p. 132). As described above, Peterson and Seligman (2004) applied exclusively theoretical approaches (i.e., no data collection based on self- or informant reports involved) to abstract six universal core virtues from religious and philosophical writings, to compile 24 (more specific) character strengths, and to map the latter onto the former in a Linnaean manner, thereby tying measurable character traits in with virtue tradition. Transferring human character to psychology as a measurable framework allows for compatibility with various other disciplines, such as ethics, pedagogy, theology, and philosophy, in which character/virtue—rather than personality—is a common concept.

Neuropsychic Conception Based on Neutral Input Versus Sociopsychic Conception Based on Highly Evaluative Input. As described above, Peterson and Seligman (2004) winnowed and condensed a comprehensive collection of positive characteristics (i.e., a high proportion of morality-relevant, thus highly evaluative input) to obtain the 24 positively, mostly also morally valued character strengths. Thereby, VIA character strengths capture what Saucier (2019) calls a “sociopsychic” or “psychocultural” conception of human beings that considers the individual in morally relevant social and cultural contexts.

By contrast, psychologists deliberately excluded morality for a long time from their conception of personality (Saucier, 2019), which dates back to Allport and Odbert’s (1936) first systematic large-scale lexical approach: While including terms they had classified as “neutral”, they excluded person-descriptors that convey social judgements of behavior in terms of its rightness/wrongness. Thereby, they attempted to capture an individualistic or “neuropsychic” conception of personality that considers intrinsic, temperamental dispositions of autonomous individuals (John et al., 1988; Saucier, 2019). Allport and Odbert’s (1936) procedure heavily influenced subsequent approaches to personality.

Saucier (2019), however, showed that different psycho-lexical personality studies conducted throughout the 20th century (i.e., Allport & Odbert, 1936; Goldberg, 1990; Norman, 1967) did not (manage to) fully exclude person-descriptors with moral content. They only did so to the degree that prevented the emergence of a separate morality-relevant factor, thus resulting in a five-factor solution (i.e., the Big Five). If a critical number of morality-relevant adjectives is included in psycho-lexical studies, a corresponding factor occurs, resulting in a six-factor solution (i.e., the Big Six; e.g., Lee & Ashton, 2008).

Saucier’s (2019) findings have the following implications for an empirical comparison of character and personality traits: First, because the Big Five are not solely based on neutral content but also—inadvertently—on some moral content (i.e., a low proportion of highly evaluative input), there likely are few nearly redundant VIA and Big Five personality traits even though a morality-relevant factor is absent in the Big Five framework. Second, given the presence of a morality-relevant factor in the Big Six framework, even more nearly redundant VIA and Big Six personality traits are to be expected. However, it is unlikely that the VIA trait space is fully subsumed under the Big Six trait space. While including some psychocultural/sociopsychic dispositions in their conception of personality in addition to the predominantly neuropsychic dispositions covered by the Big Five (Saucier, 2019), the Big Six are still not the result of an *all-inclusive* lexical approach. For example, in their study extracting the HEXACO factors based on prototypically or frequently used terms, Lee and Ashton (2008) included only 34 out of the 50 most morality-relevant adjectives identified by Saucier (2019) (i.e., an intermediate proportion of highly evaluative input). Furthermore, in the HEXACO model, both neuropsychic and sociopsychic dispositions are recovered in 24 facets, whereas the 24 VIA facets focus exclusively on sociopsychic dispositions.

Empirical Evidence. Fowers et al. (2021) summarized several studies that examined correlations of, amount of explained variance in, and incremental validity of different character constructs (justice, forgiveness, gratitude, and kindness) with/by Big Five traits supporting the distinctness between character and personality constructs under consideration. McGrath et al. (2020) examined the correlation-based overlap between the VIA character strengths on the one hand and the facets of both the FFM and the HEXACO model on the other hand. They concluded that eight out of 24 VIA character strengths (kindness, prudence, appreciation of beauty, hope, creativity, zest, forgiveness, and industriousness/perseverance) “could be considered essentially redundant” (p. 130) with a HEXACO facet, while the same is only true for appreciation of beauty and the *Aesthetics* facet from the FFM. Furthermore, McGrath et al. (2020) showed incremental validity of the 24 character strengths beyond the facets of different FFM and HEXACO inventories as predictors for different (e.g., behavioral and clinical) criteria by jointly regressing all facet scales of a personality inventory on a criterion in the first step, adding all 24 character strength scales in a second step, and evaluating ΔR^2 coefficients. Likewise, they also showed incremental validity of the different personality inventories beyond VIA inventories. To sum up, while showing some similarities or even redundancies, McGrath et al.’s (2020) findings mainly point to a certain distinctness of personality and (VIA) character and their added value beyond one another. Ruch et al. (2021) complemented these findings by showing that the NEO facets were not able to jointly explain all of the true score variance in any of the character strengths.

Measuring VIA Character Strengths

Most VIA instruments are available from the VIA Institute on Character (n.d.), where researchers can use them free of charge but only upon registration. The VIA Inventory of Strengths (VIA-IS; Peterson et al., 2005; Peterson & Seligman, 2004; VIA Institute on Character, n.d.) is the full version of the original VIA survey for adult assessment comprising 240 positively keyed items to measure the 24 character strengths. The VIA-IS was transferred to the open access International Personality Item Pool (IPIP; Goldberg et al., 2006) with minor adaptions (i.e., minimal item rewordings to fit IPIP item structure, conversion of some items to negative keying, and adjustment of the response scale to the IPIP standard). To achieve Cronbach's alpha ≥ .70 for each of the 24 scales, IPIP omitted 39 items of the VIA-IS and added 12 non-VIA items from the IPIP website, resulting in a 213-item IPIP-VIA inventory (with the 39 omitted items also published on the IPIP website, 252 items are publicly available; IPIP, n.d.).

As per documentation by McGrath (2019), the scale authors of the VIA-IS only applied one criterion in its development: The 24 character strength scales should show satisfactory Cronbach's alpha coefficients. Obviously, by only considering internal consistency, the authors of the VIA-IS did not satisfy state-of-the-art scale development and may inadvertently even have counteracted other important psychometric properties of the scales (e.g., content validity). The potential shortcomings of the VIA-IS are detailed in Study 1 of this dissertation. They also largely apply to its only minimally modified open-access cousin, the IPIP-VIA. At one glance, the IPIP-VIA scales lack:

- (Demonstration of) <u>essential unidimensionality</u> showing that each scale primarily measures its targeted character strength, while secondary (e.g., method) factors generate only small proportions of variance (Stout, 1987). Evidence by Ng et al. (2017), who deleted 133 out of 240 VIA-IS items to yield unidimensional CFA models for each scale, suggests that (essential) unidimensionality is unlikely to hold for the IPIP-VIA scales.

- Appropriate reliability estimates such as McDonald's (1999) omega coefficients based on tau-congeneric measurement models.
- Content validity. Going over the IPIP-VIA reveals different shortcomings in item construction that impair the scales' content validity. Some items miss the definitional core of their targeted strength or its skill-like nature. Furthermore, some items use idiomatic expressions, vague quantifiers, hardly observable or ambiguous content. Finally, some scales represent the underlying construct too narrowly. Study 1 presents sample items illustrating these shortcomings, some of which have also been raised by McGrath (2019), McGrath and Wallace (2021), and Noftle et al. (2011) regarding the VIA-IS.
- Cross-cultural applicability including (demonstration of) measurement invariance—despite the character strengths' claim to universality. For example, items referring to dieting to assess Self-control, to using the internet to assess Love of Learning, or to the time invested in prayer/meditation to assess Spirituality/Religiousness are unlikely to be measurement invariant across diverse cultures or not even universally applicable.
- (Full) acquiescence control. Acquiescence, that is, the tendency to agree with items irrespective of their content (Podsakoff et al., 2012), is especially problematic in cross-cultural survey research based on heterogeneous samples (Lechner et al., 2019). Effective correction of acquiescence bias requires fully balanced item keying within a scale. However, IPIP-VIA scales only partly balance positively and negatively keyed items (while VIA-IS consists of positively keyed items only).
- Economical length. Given their length of > 200 items, both IPIP-VIA and VIA-IS impose a high response burden which is unfeasible in (large-scale) surveys.

- Factorial validity. As outlined above, studies on the factor-analytic VIA trait hierarchy and its global dimensions (e.g., Brdar & Kashdan, 2010; McGrath, 2015; Ruch et al., 2010; Singh & Choubisa, 2010) yielded inconsistent results, and even the three replicable global dimensions *Caring*, *Inquisitiveness*, and *Self-Control* (e.g., McGrath, 2015; McGrath et al., 2018, 2022; McGrath & Wallace, 2021) vary considerably across studies regarding their highly loading character strengths.

These psychometric shortcomings might have compromised previous findings based on the IPIP-VIA or the VIA-IS, for example, on the character strengths' nomological net with Big Five/Big Six constructs (i.e., McGrath et al., 2020; Ruch et al., 2021). Only recently, the VIA Institute on Character revised the VIA-IS (and its short versions) and now provides the 192-item VIA-IS-R consisting of 8-item character strength scales that each include 1–4 negatively keyed items (alongside both a positively keyed and a largely balanced-keyed 96-item short version and a nested 24-item inventory to measure the three global dimensions Caring, Inquisitiveness, and Self-Control, named the VIA-IS-V3). In the development of the VIA-IS-R, McGrath (2019) addressed several of the psychometric shortcomings of the VIA-IS. However, as mentioned above, these revised scales are not available in the public domain. The open access IPIP-VIA, on the other hand, has not yet been revised and is still suffering from its "teething problems". The above list of psychometric shortcomings suggests that the IPIP-VIA can be successfully revised by selecting items that form essentially unidimensional, reliable, balanced-keyed short scales, while weeding out items that contribute to poor content validity or cross-cultural applicability.

Measuring Global Levels of the Factor-Analytic VIA Trait Hierarchy

VIA character measurement still centers on the fine-grained facet level of 24 character strengths. However, to describe good character parsimoniously and to predict its global outcomes, researchers need measures on an aggregated, global level. Moreover, large-scale surveys that cover many different constructs and can allocate only limited questionnaire space to each of them, or studies that would like to assess character as a non-focal “side construct”, might opt for parsimonious global-level character scales. Whereas the BFI-10 for an ultra-short measurement of the Big Five factors (Rammstedt & John, 2007) is included in several survey programs (e.g., in the sixth wave of the World Values Survey (WVS) and the International Social Survey Programme (ISSP) of 2005; Inglehart et al., 2018; ISSP Research Group, 2013), scales of global VIA dimensions for the use in large-scale surveys are not available yet.

Merits of the IPIP and Public Domain Instruments

Why is it important to provide measurement instruments in the public domain, such as in the IPIP? First, it helps researchers to track, replicate, and synthesize empirical findings based on these instruments—especially if authors publish their analysis code and data (i.e., open material and open data). Second, it allows and encourages the research community to collectively revise and advance these instruments (e.g., Goldberg et al., 2006). Third, survey programs such as the WVS (n.d.) only include measurement instruments that they can publish along with the data. Thus, to advance replicable character measurement and to further establish character constructs in the social sciences by paving their way into (international) survey programs, we need valid, cross-culturally applicable, and economical public domain scales.

The Present Dissertation

Research Aims: Meeting Methodological Challenges in VIA Character Assessment

Several methodological challenges in VIA character assessment arise from the above. To address them, this dissertation pursues three research aims in three sequential studies, with three general approaches across all studies: First, the dissertation follows a rigorous *open science approach*, including the revision and development of public domain character scales, the pre-registration of hypotheses, open data and material published on the Open Science Framework (OSF), and open access publications or preprints of the studies. Second, the dissertation follows a strict *(cross-country) replicability approach* by always using large, heterogeneous, and parallel samples of the German and the British population from two or three different data collections. Third, the dissertation follows an *integrative approach* to character (assessment). That is, it integrates its findings on character into the bigger picture of personality science or individual differences research. Doing so accounts for the overlapping content and conception of character and personality constructs to avoid theoretical confusion or jingle-jangle fallacies between them. It includes the construction of content-valid scales consisting of thoroughly worded items to capture conceptual/definitional differences between character and personality constructs. Moreover, the integrative approach involves relating character constructs and corresponding scales from different levels of the VIA trait hierarchy both theoretically and empirically to those from, for example, the Big Five framework (e.g., Soto & John, 2017a) and the basic human value framework (e.g., Schwartz, 1994).

The goal of Study 1 was to develop economical, content-valid, balanced-keyed (i.e., corrected for acquiescence), and cross-culturally comparable scales to measure the 24 VIA character strengths in the public domain. To this end, Study 1 aimed to develop a revised version of the IPIP-VIA, the 96-item IPIP-VIA-R, simultaneously in English and German, based on a rational-empirical approach that encompassed both expert ratings on item quality and psychometric analyses of self-report data (cf., Soto & John, 2017a). Building on Study 1, the goal of Study 2 was to revisit the hierarchical structure of the 24 VIA character strengths with a rigorous factor-analytic approach based on the IPIP-VIA-R. In doing so, Study 2 aimed to identify useful global dimensions and lay the ground for a (cross-culturally) valid VIA trait hierarchy. The goal of Study 3 was to develop economical scales to measure the three *global core strengths* identified in Study 2: *positivity*, *dependability*, and *mastery*. To this end, Study 3 aimed to select three 6-item scales from the IPIP-VIA-R that simultaneously optimize construct coverage/content validity, balanced keying, measurement model fit, reliability, and measurement invariance based on an algorithmic item selection approach.

Data Base

This dissertation is based on one ad hoc sample from the US collected in 2004 (analysis sample size: N = 713; focal variables: IPIP items, including the IPIP-VIA) and eight large and heterogenous samples from four data collections in both Germany and the UK conducted in January 2018 (analysis sample sizes: N_{Germany} = 466; N_{UK} = 455; focal variables: preliminary version of IPIP-VIA-R and validation variables), October 2018 (analysis sample sizes: N_{Germany} = 476; N_{UK} = 474; focal variables: final version of IPIP-VIA-R and validation variables), October 2019 (analysis sample sizes: N_{Germany} = 468; N_{UK} = 476; focal variables: final version of IPIP-VIA-R and validation variables), and July 2021 (analysis sample sizes: N_{Germany} = 420; N_{UK} = 440; focal variables: three 6-item core strength scales developed in Study 3 and validation variables) by a commercial online access panel provider (Respondi, 2022). For the four data collections in 2018–2021, quota samples of the German population cross-classifying gender, age cohorts, and educational attainment levels, plus parallel samples from the UK population, were drawn.

To develop and validate the 24 facet-level character strength scales, Study 1 drew on the ad hoc sample from the US and the data from the first and second data collection. To revisit the hierarchical structure of the 24 VIA character strengths, Study 2 pooled the data from the second and third data collection (i.e., separately for Germany and the UK). Study 3 drew on the samples from the second, third, and fourth data collection to train the algorithm to select items for the three global core strength scales, validate the algorithm's item selection, apply the final item selection, and validate the core strength scales based on a broad set of validation variables.

Summary of the Three Studies' Findings

Study 1: Towards Shorter, More Content-Valid, and Cross-Culturally Comparable IPIP-VIA Character Strength Scales

Along seven research questions, Study 1 examined (1) as a preparatory step for item selection, the source material (i.e., the IPIP-VIA items provided for the measurement of each character strength) by exploratory factor-analytic means, (2) if it was possible to select balanced-keyed, content-valid, and cross-culturally applicable or adaptable 4-item character strength scales from the source material (i.e., IPIP-VIA or the overall IPIP, alternatively) that would prove essentially unidimensional, (3) if these revised IPIP-VIA-R scales showed good measurement model fit, (4) if the IPIP-VIA-R scales showed good reliability, (5) if the IPIP-VIA-R scales were measurement invariant across samples from the UK and Germany using a German adaption of the IPIP-VIA-R based on the TRAPD approach (Harkness, 2003), (6) if the IPIP-VIA-R scales showed construct validity in a nomological net with basic personality and value traits, and (7) if the IPIP-VIA-R scales showed (incremental) criterion validity (beyond basic personality traits). (Sub-)Studies 1.1−1.7 addressed these research questions.

Study 1.1 underscored the need to develop short(er) character strength scales and provided empirical indications of suitable candidate items for these short scales. In a subsample of $n = 711$ from the "Eugene-Springfield Community Sample" (ESCS; Goldberg & Saucier, 2016), principal axis factoring (PAF), parallel analysis (Horn, 1965), and minimum average partial (MAP) analysis (Velicer et al., 2000) suggested that the 24 IPIP-VIA scales often lack unidimensionality. Furthermore, both low factor loadings in single-factor solutions and low item communalities in multi-factor solutions suggested that several IPIP-VIA items barely capture their targeted character strength.

In Study 1.2, four expert raters were able to identify 96 items mainly from the 252 IPIP-VIA items and additional 126 non-VIA IPIP items to constitute a balanced-keyed 4-item scale for each character strength, thereby optimizing content validity, concise wording, and cross-cultural applicability/adaptability. Intra-class correlation coefficients indicated good inter-rater reliability. Furthermore, the content validity indices of the scales (i.e., the average of the relevance scores of the four selected items derived from the expert ratings) suggested that 20 out of 24 scales had good or very good content validity, whereas for four scales the source material did not allow for converging relevance ratings. After translating/adapting the scales to German, first empirical evidence (unidimensionality, model fit, and reliability indices) in both countries based on heterogeneous samples from the UK (N = 455) and Germany (N = 466) suggested trying out five alternative items and additionally five alternative translations. A comparison of the scales with initially versus subsequently selected/translated items based on fresh heterogeneous samples from Germany (N = 476) and the UK (N = 474) yielded the final IPIP-VIA-R scales. Two unidimensionality indices (ECV, explained common variance, and MIREAL, mean of item residual absolute loadings; Ferrando & Lorenzo-Seva, 2018) suggested that, overall, the final IPIP-VIA-R scales achieved essential unidimensionality in both countries with slightly better results in the German sample.

In Study 1.3, most CFA models of the 24 IPIP-VIA-R scales, in which the four items were always regressed on one latent character strength variable and one manifest covariate capturing acquiescence (i.e., unidimensional measurement models plus a method factor), showed good fit evaluated based on χ^2 statistics and resulting goodness-of-fit indices (CFI, RMSEA, SRMR). Only the scales measuring spirituality, hope, and leadership showed insufficient model fit, suggesting that—at least within the IPIP—these strengths cannot be measured with essentially unidimensional scales because of secondary substantive factor(s).

Study 1.4 showed that—measured against the relatively short length of the scales—the reliability of the manifest scale scores was mainly but not always sufficient (judged by both the size of reliability coefficients and their comparison with those of established personality scales of the same length). McDonald's (1999) omega as most suitable coefficient of internal consistency reliability exceeded .70 (.65) for 22 (23) scales in the UK sample and for 18 (23) scales in the German sample. Test-retest correlations based on a time lag of about three weeks suggested sufficient temporal consistency/stability of most manifest scale scores in Germany (r_{tt} > .70 (.65) for 19 (21) scales), while the results in the UK sample were somewhat lower (r_{tt} > .70 (.65) for only 8 (14) scales). These findings generally greenlight the use of manifest scale scores, while suggesting the use of latent measurement models if feasible.

Measurement invariance analyses of Study 1.5 yielded encouraging evidence for the cross-cultural applicability of the IPIP-VIA-R. Comparing multi-group CFA models of different measurement invariance levels based on changes in goodness-of-fit indices (CFI, RMSEA, and SRMR) and the BIC information criterion showed (partial) scalar invariance of all IPIP-VIA-R scales between the UK and Germany, suggesting that covariance-based statistics and latent means of all character strengths can be compared across countries.

Locating the 24 IPIP-VIA-R character strengths in a nomological net with Big Five domains and facets and basic human value orientations in Study 1.6 yielded clear, value-laden, and theory-conforming validity patterns consistently across the UK and Germany. Bivariate Pearson correlations between IPIP-VIA-R scale scores and both Big Five domain scores (measured with the BFI-2-S; Soto & John, 2017b) and the four basic human value cluster scores (measured with the PVQ-21; Schwartz, 2003a) provided first evidence for convergent validity by often confirming our hypothesis about the strongest correlation between a character strength and a Big Five domain or a value cluster. At the same time, nearly all strengths substantially correlated with all Big Five domains, showing that character strengths usually reside at the intersection of all Big Five personality traits. Nevertheless, disattenuated correlations and multiple regressions, regressing each character strength on all Big Five domains or facets, suggested that most character strengths are not redundant to Big Five traits. Locating the character strengths in the two-dimensional value orientation space (axis 1: self-transcendence vs. self-enhancement; axis 2: openness to change vs. conservation) again showed a theory-conforming pattern: All strengths (but leadership in Germany) were associated with an orientation towards self-transcendence, that is a motivation to renounce self-interested actions—some of them in combination with an orientation towards openness to change, that is a motivation to personally grow, and some of them in combination with an orientation towards conservation, that is a motivation to preserve social group functioning.

Study 1.7 yielded initial evidence for the criterion validity of IPIP-VIA-R and its incremental validity beyond the Big Five based on two relevant outcomes of the “good life”, namely, life satisfaction and health. The evidence tended to be stronger in the German sample than in the UK sample. Bivariate Pearson correlations between each character strength and both outcomes were in line with theoretical expectations or replicated previously reported patterns (e.g., positive association between most strengths and life satisfaction). Hierarchical regression analysis showed that not all but some character strengths explained additional variance in life satisfaction or health beyond the five Big Five domains—sometimes as much as each Big Five domain on average—or beyond the 15 facets.

To sum up, Study 1 made a new contribution to the public domain toolbox of personality science that allows to assess human character economically, reliably, validly, and comparably across the UK and Germany in survey-based research. Furthermore, by locating character strengths—as measured by the newly developed IPIP-VIA-R scales—in the trait space of basic personality traits and human value orientations, by looking at their associations with relevant outcomes, and by examining their additional predictive capacities, Study 1 both replicated and complemented previous findings and provided new insights into the role and added value of character traits in personality science. Even though few minor issues remain with some IPIP-VIA-R scales (that might be solved by future research by dispensing with the self-imposed restriction to extant IPIP items), translation or adaption of the scales to languages from more diverse cultures is in demand to more strictly test the IPIP-VIA-R’s cross-cultural applicability (i.e., one of its main features), and more analyses to further corroborate each scale's external validity (e.g., with more specific variables) are pending, the IPIP-VIA-R can be applied promptly by personality or character researchers.

Study 2: Revisiting the Hierarchical Structure of the 24 VIA Character Strengths

Study 2 introduced three criteria of *useful* global dimensions of a trait hierarchy and applied three complementary methods to reveal these dimensions of the VIA trait hierarchy. The criteria of useful global levels of abstraction were *interpretability*, *globality/generality*, and *cross-cultural replicability* of all extracted dimensions (i.e., principal components) of a given level. To evaluate the usefulness of higher-level dimensions, Study 2 distinguished between highly loading *marker strengths* (i.e., $\lambda \geq .50$) and *co-defining strengths* of medium loading size (i.e., $.30 \leq |\lambda| < .50$). The set of factor-analytic methods encompassed (a) parallel analysis (Horn, 1965) and MAP analysis (Velicer et al., 2000) to determine the number of higher-level dimensions that parsimoniously summarize the (co-) variation of the 24 character strengths in each country, (b) Bass-ackwards analysis (Goldberg, 2006) to reveal the top eight levels of the character strengths' solutions-hierarchy in each country (i.e., principal component analyses (PCAs) extracting an increasing number of components, starting with one component, then two components, up to eight components), and (c) cross-country congruency analysis (Tucker, 1951) to identify the most similar and thus replicable higher-level dimensions across countries.

Based on two large and heterogeneous, parallel samples from Germany and the UK (total $N \approx 2{,}000$), all three complementary factor-analytic methods jointly suggested that Level III of the VIA solutions-hierarchies best fulfilled the criteria of usefulness: Each of the three dimensions was characterized by a unique set (i.e., good interpretability) of at least three highly loading marker strengths (i.e., globality/generality) that replicated across Germany and the UK (i.e., cross-cultural replicability). The first dimension *positivity* was characterized by the marker strengths of forgiveness, zest, hope, and capacity for love; the second dimension *dependability* by prudence, modesty/humility, integrity, and equity; and the third dimension *mastery* by judgment, originality, perspective, valor, leadership, and social intelligence (see also Table 3.1 of Study 3). Interestingly, these IPIP-VIA-R-based global VIA dimensions were not fully congruent with those recently established by factor-analytic means based on different proprietary VIA instruments from the VIA Institute on Character (n.d.) labeled Caring, Inquisitiveness, and Self-Control (e.g., McGrath et al., 2018).

The Levels I and II of the VIA solutions-hierarchies also fulfilled the criteria of usefulness, whereby Level II lends itself for a direct comparison with personality metatraits from Level II of the Big Five and Big Six frameworks. Looking at the content overlap of the constructs, the VIA Level-II-dimensions resembled the “Big Two” from the Big Six framework (i.e., Dynamism and Social Self-Regulation; e.g., Saucier et al., 2014) more closely than the metatraits from the Big Five framework (i.e., Plasticity and Stability; e.g., DeYoung, 2006). To sum up, Study 2 contributed to establishing the VIA trait hierarchy by corroborating recent findings that the VIA trait space is spanned by three global dimensions and allowed for a theoretical comparison of global levels of VIA and personality frameworks.

Study 3: Measuring Global Character Dimensions

It has been the prevailing approach in personality measurement to measure global constructs with a subset of items that measure the subordinate, more specific constructs (e.g., Thalmayer & Saucier, 2014, who measured the Big Two metatraits with an item subset of their Big Six scales). Hence, developing three core strength scales based on the 24 character strength scales was an item selection task. Recommendations from the literature (Soto & John, 2019) and common practice in personality measurement (e.g., Entringer et al., 2021; Soto & John, 2017b; Thalmayer & Saucier, 2014) suggested the selection of six items per core strength to cover each of their 4–6 marker strengths in sufficiently reliable and valid, but at the same time economical and essentially unidimensional scales. Selecting 18 out of 96 IPIP-VIA-R items while optimizing multiple psychometric criteria (i.e., construct coverage/content validity, balanced item keying, measurement model fit, reliability, and scalar measurement invariance) was a complex combinatorial problem. This was approached with an algorithmic item selection procedure called Ant Colony Optimization (ACO). Across several iterations, the ACO algorithm selects different item sets to measure the targeted constructs and evaluates them based on multiple user-defined psychometric criteria, thereby learning which items contribute to a (close-to-) optimal item selection. ACO stops and presents its best item selection when it yields no further improvement in the psychometric criteria (Olaru et al., 2019; Schroeders et al., 2016).

Based on two large and heterogenous samples from Germany and the UK (total N = 950, the *training samples*), ACO selected 18 items that fit a multi-group correlated first-order factor model with three factors and a covariate capturing acquiescence variance. Given the user-defined sampling restrictions to ensure construct coverage and content validity, ACO selected one or two items of each marker strength to measure the respective core strengths. Furthermore, the algorithm selected two items of co-defining strengths. Study 2 showed that co-defining strengths also contributed to the makeup of the core strengths, which is why ACO could also select items from co-defining strengths. The final three core strength scales that ACO presented were each fully balanced, achieved scalar measurement invariance across countries, and showed satisfactory composite reliability (i.e., McDonald's (1999) omega coefficients) and test-retest correlations based on a test-retest interval of 2–3 weeks. Study 3 replicated these psychometric findings in four further samples from Germany and the UK. The *validation samples* (total N = 944) showed that the psychometric properties of the item selection—optimized based on the training samples—could be generalized to other samples to ensure that ACO did not overfit its selection (see Olaru et al., 2019). The *use-case samples* (total N = 860), to which only the three core strength scales were presented, showed that the scales maintained their psychometric properties when used as independent 18-item inventory.

To validate the three core strength scales, Study 3 first examined how they were located in the nomological net with personality metatraits (*Stability* and *Plasticity* from the

Big Five framework; DeYoung, 2006) and value metatraits (from the basic human values framework: *Social Growth Orientation* computed as self-transcendence minus self-enhancement and *Personal Growth Orientation* computed as openness to change minus conservation; Rudnev et al., 2016; Schwartz, 1992, 2012), because as positively valued personality traits, traits also from the global level of the VIA trait hierarchy were expected to reside at the intersection of personality and value trait spaces. Second, Study 3 examined associations of the three core strength scales with the broad criterion of general life satisfaction (Diener, 1984) and a total of 22 value-driven behaviors derived from the ten basic human values (e.g., Schwartz, 1994). All validation scales were selected on theoretical grounds, and validation hypotheses were pre-registered accordingly. Overall, the distinct association pattern of each core strength scale with the various validation scales supported the scales' convergent, discriminant, and criterion validity in both Germany and the UK. With the maximum correlation amounting to $r = .77$ (between positivity and Stability in the UK), the scale validation also suggested that core strengths are related with but at the same time sufficiently distinct from personality metatraits.

To conclude, the comprehensive scale development and validation approach of the dissertation's third study brought forth three both economical and valid public domain core strength scales. Even though further research on the scales is advisable (e.g., adapting the scales to and testing them in a more diverse set of countries or replicating some of their psychometric properties based on multi-informant ratings), Study 3 suggested that they can be readily used in survey-based character research or integrative personality and value research.

STUDY 1

Abstract

The “Values in Action” (VIA) framework is currently the most prominent approach to conceptualizing human character. VIA posits 24 character strengths that are purportedly valued across cultures and promote the well-being of both individuals and communities. However, unresolved limitations in the assessment of these character strengths continue to hamper theoretical progress in research on human character based on the VIA framework. Here we sought to lay a new foundation for advanced assessment of strengths by refining and extensively validating an existing open-science inventory from the International Personality Item Pool (IPIP). Using data from a US community sample and four quota samples from the United Kingdom and Germany, we investigated whether valid, cross-culturally comparable, and economical assessment of the VIA character strengths is possible with the IPIP. Experts selected suitable items with the aim to obtain 24 balanced-keyed short scales. Different experts then translated these items to German. Through exploratory and confirmatory factor analysis, we established essential unidimensionality and well-fitting measurement models for each scale. All scales achieved at least partial scalar invariance across languages. Reliability estimates were satisfactory. Extensive analyses of the strengths’ nomological network placed character strengths between Big Five personality traits and basic human values, confirming that VIA strengths emphasize Self-Transcendence rather than Self-Enhancement. With few exceptions, the 24 character strengths scales were sufficiently distinct from the Big Five, and many showed incremental predictive validity, also for “good life” criteria. The 96-item inventory “IPIP-VIA-R” offers a sound and fully open-science approach to future research on character strengths.[*]

Keywords: Values in Action (VIA), character strengths, Big Five traits, facets, human values

[*] A previous version of Study 1 was informally published as Bluemke, M., Partsch, M. V., Saucier, G., & Lechner, C. M. (2021). *Human character in the IPIP: Towards shorter, more content-valid, and cross-culturally comparable IPIP-VIA character strength scales*. PsyArXiv. https://doi.org/10.31234/osf.io/k79qf. I made several minor revisions to this version for inclusion in this dissertation.

Theoretical Background

Peterson and Seligman (2004) introduced the “Values in Action” (or VIA) framework and the accompanying “manual of sanities.” Their intention was to broaden psychology’s exclusive focus on psychopathology and undesirable human behavior (as expressed in the *Diagnostic and Statistical Manual of Mental Disorders*, e.g., American Psychiatric Association, 1994). With the aim to establish whether there are ubiquitous and universal virtues, and what exactly constitutes “good character”, they reviewed the work of religious thinkers and (moral) philosophers from across the Eastern and the Western hemisphere and from ancient and modern times, for example, Plato’s cardinal virtues and Aristotle’s and Augustine’s views building on them, and Aquinas’s medieval theological virtues in line with Paul’s ancient biblical writings. They also considered Eastern philosophy (Confucianism, Taoism). From this review, they derived six virtues recurring throughout time and space (the “High Six” core virtues: *justice, wisdom, temperance, courage, humanity,* and *transcendence*), which themselves are too abstract to be measured directly. Therefore, they also identified 24 potentially universal, theoretically justified, and empirically supported character strengths that feed into these core virtues. Forming the heart of measurement in the VIA framework, these 24 character strengths are also reflected in the Values in Action Inventory of Strengths (VIA-IS) (Peterson et al., 2005).

In a multi-stage process, and inspired by psychological research on positive traits, Peterson and Seligman (2004) defined and classified 24 cross-culturally valued character strengths (see Table 1.1 for working definitions). They conceptualized character strengths as relatively stable, universally recognized and valued individual differences. Within the VIA classification, character strengths belonging to the same virtue have a common denominator, yet they are not replicas of one another (or necessarily highly correlated traits) but distinctive. High standing on one strength does not imply a similar standing on another strength mapped onto the same virtue. On the contrary, displaying all the strengths classified under a virtue is nearly impossible for individuals as strengths are different ways of expressing virtues.

To be included in the VIA classification, character strengths had to satisfy most of ten features: A distinctive and trait-like character strength should be fulfilling and contribute to a good life for oneself and others via manifold pathways. It should be (morally) valued in its own right, irrespective of its outcomes, and displaying a strength should elevate, not diminish or belittle other people. Therefore, society provides institutions and rituals associated with the cultivation of character strengths. A character strength implies that no desirable opposite trait exists (nonfelicitous opposite), and the absence of a strength should be observable in some individuals (selectivity). Cognitive abilities were excluded. Peterson and Seligman (2004) did not deem the VIA classification exhaustive; instead, they invited research that might lead to a revision (even expansion) of measures and their classification, though over the past 20 years the classification has not undergone revision, attesting to a rather consolidated and well-accepted framework (Park, 2018; Peterson & Park, 2009; Ruch & Proyer, 2015).

Table 1.1

VIA Classification of Character Strengths and Associated Virtue Clusters

IPIP-VIA Scale Label	Definition of Strength	Virtue Cluster
Appreciation of Beauty & Excellence [APP]	Recognizing and valuing beauty in the physical environment, skill/talent, or virtues	Transcendence
(Capacity for) Love [CAP]	Ability to express warmth and receive love from others, valuing reciprocal caring and sharing	Humanity
Citizenship/Teamwork [CIT]	Supporting a group to which one is loyal and contributing to its cause by collaborating and cooperating effectively	Justice
Curiosity [CUR]	Desire to explore the world and seek out exciting experiences and novel ideas	Wisdom
Equity [EQU]	Treating people equally, with respect and in a fair and unbiased manner	Justice
Forgiveness [FOR]	Readily accepting the shortcomings of others and showing mercy	Temperance
Gratitude [GRA]	Recognizing good things in life, feeling appreciation, and expressing thanks	Transcendence
Hope [HOP]	Seeing a positive future and developing a way to achieve desired events and outcomes	Transcendence
Humor [HUM]	Enjoying laughter and bringing smiles to others by highlighting the funny or lighter side of life	Transcendence
Industry/Perseverance [IND]	Steadfastness in pursuing goals even in the face of obstacles, [taking pleasure in] accomplishing tasks	Courage
Integrity/Honesty [INT]	Being truthful to others and to oneself, trustworthy, and acting with moral integrity	Courage
Judgment [JUD]	Weighing all available evidence fairly and being willing to change one's opinion accordingly	Wisdom
Kindness [KIN]	Being concerned about others, helping generously without expecting reciprocity	Humanity
Leadership [LEA]	Organizing collective success of a group and fostering good working relationships among members	Justice
(Love of) Learning [LOV]	[Having joy when] expanding existing knowledge and learning new skills	Wisdom
Modesty [MOD]	Being modest (though realistic) about one's abilities and weaknesses, valuing contributions by others	Temperance
Originality/Creativity [ORI]	Having original ideas, coming up with innovative and productive ways to do things	Wisdom
Perspective [PER]	Being able to see the big picture and in a position to give good advice on essential questions in life	Wisdom
Prudence [PRU]	Being smart and careful about choices in the interest of avoiding undue risks and preventing regrets	Temperance
Self-Regulation [SEL]	Regulating emotions, thoughts, and behaviors to live up to stringent (self-imposed) standards of self-control	Temperance
Social Intelligence [SOC]	Understanding own and other's feelings and motives, knowing how to adapt socially to situations	Humanity
Spirituality/Religiousness [SPI]	Seeing a higher purpose and meaning in life, feeling connected to the sacred or believing in transcendent powers	Transcendence
Valor/Bravery [VAL]	Conquering fear and standing up for one's convictions in defiance of well-known risks	Courage
Zest/Vitality [ZES]	Approaching activities with physical and mental vitality, feeling alive and energetic	Courage

Note. IPIP-VIA = Scale labels; definitions adapted from Peterson & Seligman (2004) and Watkins (2016); Virtue Cluster = Strength association with one of six VIA virtue clusters.

Relevance of the VIA Character Framework

Character strengths as defined by Peterson and Seligman (2004) relate traits to intrinsic motivations for behavior. The notion that character strengths are *values* in action highlights that these character strengths are value-laden traits. At the same time, the idea that they are values *in action* highlights that character strengths are *skills* that are expressed in behavior and, in principle, malleable and evolving over time. Skills need to be practiced by individuals across a variety of situations to become recognizable strengths and form signature strengths. Precisely such skillful mastery is likely to support individual well-being (Seligman, 2011).

Character strengths reside at a similar level of abstraction as personality facets in the personality trait hierarchy. They form the building blocks of the VIA character trait space (McGrath et al., 2020; Partsch et al., 2022), resembling facets of traits in other trait spaces, such as the Big Five or HEXACO domains (Ashton & Lee, 2007; McCrae & John, 1992; Thalmayer & Saucier, 2014). Yet, despite these similarities, differences between character strengths and personality traits are evident. Reminiscent of Allport's (1927) distinction between devaluated and evaluated personality, VIA includes purposively selected, cross-culturally morally *valued* character traits that are not sufficiently represented in (mostly devaluated) lexical personality frameworks (e.g., Spirituality, Humor, Valor/Bravery, Social Intelligence; see Saucier, 2009).

Within personality science, the VIA character strengths offer a unique growth-oriented positive-psychological perspective that can complement other personality descriptors such as temperament, attitudes, values, beliefs, interests, or goals (Kandler et al., 2014). VIA's theoretical footing and its deliberate focus on *value-laden* and morally valued traits makes it a welcome addendum to lexical approaches in which moral and strongly evaluative person descriptors were eliminated (Peterson et al., 2005; Saucier & Iurino, 2020; Stahlmann & Ruch, 2020). Moreover, the specific appeal of the VIA framework lies in its cross-cultural inspiration. Due to its broad trans-philosophical, cross-cultural and diachronic perspectives, the VIA framework is meant to be a universally applicable individual differences model (Dahlsgaard et al., 2005). It is therefore not surprising that the framework has received roughly 10,000 citations at the time of this writing (Ackerman et al., 2018).

Criticism of Character Strength Assessment

Ultimately the viability of the VIA framework as a conceptualization of human character and its utility for applied and research purposes hinges crucially on the extent to which the 24 proposed character strengths can be validly assessed. If scales do not reliably and validly measure what they purport to measure, then analyzing relations between character strengths and life outcomes, or group differences and intervention effects, is premature (e.g., Biswas-Diener, 2006; Park et al., 2004; Park & Peterson, 2010; Seligman et al., 2005). However, despite VIA's important resource orientation and obvious relevance for clinical, educational, and workplace settings as well as its already widespread application in research and practice, the psychometric foundations of attendant measures lag far behind other individual differences constructs such as lexicon-based personality traits or values. Most of the positive-psychological literature is tellingly silent about the psychometric foundations of measuring character strengths (but rich in delving into philosophical perspectives and inspecting consequential outcomes). Especially in the public domain, there are persisting limitations in the measurement of character strengths, including few validation attempts that limit progress in the field and wider acceptance of character strengths in standard personality testing.

The authors of the VIA framework initially developed the proprietary 240-item VIA Inventory of Strengths based on self-report (VIA-IS for adults; Peterson et al., 2005; Peterson & Seligman, 2004; VIA Institute on Character, n.d.). In this research, we use public-domain IPIP-VIA items from the International Personality Item Pool (IPIP; Goldberg et al., 2006). The IPIP-VIA is based on the VIA-IS (provided by Peterson in June 2001 and described in

Peterson and Seligman (2003)), was slightly adjusted to enter IPIP and underwent a first revision in 2005 (see below). Since then, several abbreviated versions (e.g., VIA-IS-120), target-group specific versions (e.g., VIA-Youth), and recently also revised proprietary versions (e.g., VIA-IS-R) have been developed (McGrath, 2019). There are also modified and alternative versions such as the Global Assessment of Character Strengths (McGrath, 2019), a 24-vignettes based Character Strengths Rating Form (Ruch et al., 2014), and the structured interview focused on signature strengths (Peterson, 2003), but they are not at focus in this research.

The quality of the VIA-IS, that dominated VIA character research until recently, has been repeatedly criticized. This criticism largely applies also to the open access cousin of the VIA-IS, the IPIP-VIA, which is the focus of the present research. For example, Ng et al. (2017) found that more than half of all items from the VIA-IS needed to be excluded to achieve well-fitting measurement models for the 24 strengths (an iterative procedure merely based on low factor loadings). McGrath (2019) and McGrath and Wallace (2021) highlighted that many items lacked content validity and were too situation-specific to qualify as character (i.e., trait) measures. Even when good fit was found for measurement models for the 24 strength scales (Feraco et al., 2022), it pertained to the *truncated* (Italian) version of the VIA-IS-120—a tool for which not only half of the items were dropped simply on the basis of maximizing item-total correlations, but that exclusively consists of positively-keyed items—and does not generalize to the predominantly used instruments. Besides these psychometric limitations, Noftle et al. (2011) called for addressing construct validity of VIA scales by investigating links between strength scales and Big Five scales including their facets. The few recent studies addressing this request are based on the IPIP-VIA and both full and abbreviated versions of the original VIA-IS (McGrath et al., 2020; Ruch et al., 2021) and therefore need to be replicated with revised, psychometrically sound instruments. Below, we discuss these various points of criticism in detail and raise several additional ones.

Improper Reliability Estimates. Both scale development of the VIA-IS (Peterson et al., 2005) and scale revisions in the IPIP-VIA focused mainly on achieving high levels of Cronbach's alpha coefficients for the scales (Goldberg et al., 2006; McGrath, 2019; Peterson & Seligman, 2004; VIA Institute on Character, n.d.). Unfortunately, alpha is a less-than-optimal criterion for scale development: It yields biased reliability estimates unless the scale satisfies a tau-equivalent measurement model (McNeish, 2018; Raykov, 1998, 2001), a requirement that these authors did not test and that neither the focal IPIP-VIA nor the VIA-IS are likely to meet. Although researchers regularly analyze composite scale scores for the 24 character strengths (e.g., McGrath, 2014, 2015, 2016), the reliability of such composites has not been established by fitting adequate latent measurement models for a long time. This started to change with the recent introduction of the VIA-IS-R (see McGrath, 2022; McGrath et al., 2022; Vylobkova et al., 2023). However, there is still a need for public domain strength scales that rest on proper measurement models and provide more trustworthy internal consistency estimates. These need to be complemented by estimates of test–retest reliability to provide more direct evidence of the replicability of scale scores (McGrath & Wallace, 2021, provide such estimates of test-retest reliability for the VIA-IS-R).

Lack of Unidimensionality. Unidimensionality refers to the quality of measuring exclusively a single construct, that is, the (trait) concept of interest (Hattie, 1985). Multidimensionality implies that more than one dimension underlies the items, rendering measurement of the construct complex. A realistic goal is *essential* unidimensionality, which reflects a dominant dimension while the conditional covariances over all item pairs, on average, must be small in magnitude (Stout, 1987). However, in practice, even "satisfactory" internal consistency coefficients are often misinterpreted as evidence for scale unidimensionality (Clark & Watson, 2019; Cortina, 1993). Being largely silent about the requirements of unidimensionality of content-valid items was a questionable starting point for various studies examining factor structures or analyzing reliability of character strength scales. For example, despite the lack of suitable measurement models in general, several major studies have already begun to address the *higher-order structure* of character strengths (e.g., McGrath, 2014, with $N > 450{,}000$; see also Ng et al., 2017). However, when shedding light on the factorial validity of strength assessment, factoring whole item pools with multidimensional and interrelated strengths in the presence of overlapping items is unsuitable and likely to result in findings that do not replicate (see Feraco et al., 2022).

Only recently, based on a review of 19 studies, Feraco and colleagues (2022) complained about the lack of evidence for unidimensional character assessment in the VIA-IS tradition, aside from the general paucity of testing proper measurement models for each character strength. In their own study, they comprehensively assessed the psychometric properties (i.e., unidimensionality, measurement models, hierarchical structure, and criterion validity) of the Italian version of the VIA-IS and found generally encouraging results for the short 120-item version, despite remaining issues with some dimensions (e.g., Love of Learning). On the other hand, Ng et al.'s (2017) findings evidenced that the scales of the full version of the VIA-IS generally lack unidimensionality suggesting the same for the open access IPIP-VIA, for which the dimensionality of scales has not been examined yet. Recently, McGrath et al. (2022) showed that unidimensional measurement models did not fit for the majority of the VIA-IS-R and VIA-IS-M scales (i.e., an abbreviated version of the VIA-IS-R), which, however, they mainly attributed to method variance. From a different (somewhat less rigorous) approach to evaluate unidimensionality, McGrath (2022) concluded that the majority of the VIA-IS-R and VIA-IS-M scales were unidimensional after all.

Despite few studies having recently analyzed the dimensionality of different VIA character strengths scales, there is not yet enough awareness that factor analyses of each strength scale are required as a first step to determine their dimensionality (see Ferrando & Lorenzo-Seva, 2018). Only after the strength measurements have been optimized locally can their mutual relationships and hierarchical structures as well as their wider nomological net be validly studied (Boateng et al., 2018).

Lack of Content Validity and Deficiencies in Item Wording. Content validity is arguably the most important property of a scale (Clark & Watson, 2019). Maximizing internal consistency not only disregards content validity but can even conflict with it. A close look at some IPIP-VIA items highlights several issues with content validity that were likely present in VIA inventories from the initial development phase. First, quoting from IPIP-VIA, it is evident that not all items capture the definitional core of a character strength well despite high internal consistency. For example, "[I] have an imagination that stretches beyond that of my friends" (Originality) might not reflect useful creative imagination but escapism and daydreaming. Similarly, "[I] helped a neighbor in the last month" (Kindness) is probably too narrow and situation-specific.

Second, some items do not conform to the definition of character strengths as skills. Whereas some items do contain skill-related wording such as "being able to" or "being good at", many do not and instead represent mere attitudes (e.g., Citizenship/Teamwork: "[I] don't *think it's important* to socialize with others"; Equity: "[I] *believe* that everyone should have a say") [emphasis added].

Third, several items present content that is difficult to observe, ambiguous, vaguely quantified or idiomatic, such as "[I] read all the time" (Love of Learning), "[I] behave in unusual and strange ways" (Prudence), "[I] know what makes others tick" (Social Intelligence), "[I] prefer to participate fully rather than view life from the sidelines" (Zest), and "[I] don't approach things halfheartedly" (Zest). Ambiguous concepts, vague quantifiers, and idioms are either susceptible to subjective interpretations and range-restriction (Wänke, 2002), or difficult to translate and to adapt to other cultural contexts, thereby risking confusion among participants (Adelnia & Vahid Dastjerdi, 2011; Liu, 2012). Interpretational ambiguity also occurs when items are not distinct but overlap with other strengths. For instance, “[I] refuse to take credit for work I have not done” (Equity) and “[I] take pride in not exaggerating who or what I am” (Integrity) convey Modesty in parallel. As others have noted before (e.g., Noftle et al., 2011), some items are not only hard to objectify but even difficult to evaluate subjectively, such as “[I] have never given bad advice to a friend” (Perspective) or “[I] am trusted to keep secrets” (Integrity), thereby inviting biased evaluations (John & Robins, 1993; Vazire, 2010).

Finally, some scales provide imbalanced representations of the respective character strengths. For example, selecting Self-Regulation items with highly specific content, such as “[I] can stay on a diet” and “[I] can’t resist eating candies or cookies if they are around“ creates a *diet*-specific facet in this measure and omits other domains in which Self-Regulation may manifest. Apart from this, the items presuppose that food is readily available and that holding a diet for health reasons is a widespread practice—a situation often found in economically developed (WEIRD) countries but less so in the Majority World.

Lack of Cross-Cultural Applicability. The latter issue with Self-Regulation highlights another problem: Although Peterson and Seligman (2004) selected character strengths that they deemed cross-cultural universals, existing VIA scales often use person-descriptors that are not applicable across cultures. Using items that refer to highly culture-specific concepts hampers cross-cultural comparability (Allalouf et al., 1999). For example, like the “diet” items cited above, the Spirituality/Religiousness item “[I] have spent at least 30 minutes in the last 24 hours in prayer or meditation” is likely to function differently across cultures where institutionalized prayers or meditation are common or not (e.g., Muslims’ five daily prayers, or Salat). Similarly, “[I] go out of my way to attend educational events” (Love of Learning) will be understood differently in countries with free access to education compared to countries in which education strongly depends on socioeconomic status. Items that presuppose cultural codes and access to technology are likely prone to cultural bias, for example, “[I] consult the library or the Internet immediately if I want to know something”

(Love of Learning). Character assessment that quantifies strengths and compares countries in terms of their scale scores implicitly subscribes to a generalizing (“etic” as opposed to culture-specific “emic”) approach (Berry, 1969, 1999). However, we are unaware of any previous attempts at ensuring broad cross-cultural applicability of character strength items during item selection and testing measurement invariance of each strength scale across different languages, countries, or cultures. Before the equivalence of strength measurement has been firmly established, any statistical comparisons across countries are questionable (as in Biswas-Diener, 2006; Park et al., 2006).

Acquiescence. Acquiescence (i.e., the tendency to agree with survey questions irrespective of item content and keying, or “yeah-saying”) is a widespread source of bias in survey research (Podsakoff et al., 2012). This bias is especially pressing in research involving samples that differ in their levels of acquiescent responding, as is regularly the case in cross-cultural research and in research involving respondents with varying levels of education and ability (for a review, see Lechner et al., 2019). If unaccounted for, acquiescence can introduce substantial bias in the means, factor structures, measurement invariance tests, and validity coefficients of personality and character strength scales (Lechner & Rammstedt, 2015; McCrae et al., 2001; Primi et al., 2020; Rammstedt & Farmer, 2013; Soto & John, 2017a, 2017b). Acquiescence is likely to diminish correlations between semantically opposite item-pairs and create a method factor among the reverse-keyed items (Cambré et al., 2002; Kam & Meyer, 2015). Using balanced-keyed scales and modeling acquiescence as an additional factor is an established means of eliminating this bias in personality scales (Aichholzer, 2015; Billiet & McClendon, 2000; Soto & John, 2019). Current scale development practices clearly suggest that controlling acquiescence can largely eliminate the abovementioned biases. Thus, adequate measurement of character strengths requires balanced-keyed items that can eliminate acquiescent response bias. However, most empirical research to date has relied on unbalanced

VIA-IS or IPIP-VIA character strength measures with only positively-keyed items or an unequal number of positively and negatively keyed items.

Scale Length. Any 240-item inventory entails substantial respondent burden that can impede the quality of measurement. When social surveys of the general population are intended, or large-scale assessment takes place in international contexts, lengthy assessments are rarely feasible. Short forms alleviate respondent burden (see Rammstedt & Beierlein, 2014; Schroeders et al., 2016; Ziegler et al., 2014) while often retaining much of their longer cousins' reliability and predictive validity (Rammstedt et al., 2021; Thalmayer et al., 2011). One challenge of short scales is retaining their validity via their breadth of content. The sweet spot found for Big Five scale length was around six items for broad domain scores (as in the BFI-2-S; Soto & John, 2017b); for narrower facets, as few as two items can typically suffice (as in the BFI-2-S), although four items are advantageous (as is the case in the full BFI-2; Soto & John, 2017a). Combined with the goal of balancing each scale (acquiescence control), the challenges are obvious but—as the Big Five show—can be overcome.

The Present Investigation

Rationale and Objectives

Progress in the study of human character will require high-quality measures of character strengths that are comparable across languages and cultures. Both psychometric quality and cross-cultural validity must be considered early during scale development (Danner et al., 2016; Fischer & Poortinga, 2018; Ziegler, 2014; Ziegler & Bensch, 2013). Because—for most of the time since the introduction of the VIA character strengths—relatively little effort has been invested in constructing and validating scales that satisfy the requirements we discussed above, the already large body of evidence on character strengths does not rest on a solid psychometric footing (e.g., Martínez-Martí & Ruch, 2017). Only once such measures are available can one begin to answer fundamental theoretical questions about the stability of character traits, their relations to other individual difference constructs such as personality traits and values, or their predictive power for life outcomes.

Here we set out to advance the assessment of human character. Our aim was to establish whether it is possible to construct refined measures of the 24 VIA character that are (1) short scales (four items per strength, i.e., 96 items in total) and hence suitable for general population surveys, (2) content-valid, (3) cross-culturally applicable, and (4) balanced-keyed to control for acquiescent responding. Our additional self-imposed constraint was to (5) exclusively use existing items from the public domain, more specifically, those from the IPIP (Goldberg, 1999), which is freely available to researchers in line with open-science standards (Ashton, 2005).

The challenge was thus to achieve good psychometric properties while keeping scale length to a minimum. Short scales are sometimes criticized on the grounds that they may not be able to adequately represent a construct and achieve sufficient reliabilities (John & Soto, 2007). However, these concerns are less pressing when—as in our present case—assessing narrow facets that can be well represented with a few items. Moreover, psychometric research generally allays many of the concerns about short scales. Decreasing the number of items under unidimensionality assumptions does *not* affect criterion validity, nor does it shift latent population means (Heene et al., 2014). Moreover, the super-brief BFI-10 can predict most outcomes as well as longer Big Five scales with only two items per Big Five dimension; the BFI-2-XS with three items per dimension performs slightly better and at a level comparable to much longer personality scales (Rammstedt et al., 2021; Thalmayer et al., 2011). Given that three items can suffice to represent broad personality domains, it should be absolutely sufficient and a good compromise between bandwidth and fidelity to measure narrow facets with four-item scales (e.g., Rammstedt et al., 2021; Soto & John, 2017b; Thalmayer et al., 2011). Especially when using latent-variable models, the lower reliability of short scales does not impair predictive validity. Thus, the benefits of short scales in terms of enabling comprehensive assessments with minimal respondent burden arguably outweigh their disadvantages for our present intent, which was to create refined scales for research purposes (not individual diagnostics).

Study Overview and Research Questions

To revise and refine the IPIP-VIA character strength scales, we proceeded in five steps. First, we scrutinized the items of the IPIP-VIA and reexamined known issues with unidimensionality and convergence of items of each strength on their respective common factor (i.e., their common variance proportion and factor loadings, based on a US community sample). Second, experts evaluated the IPIP items for their closeness to the VIA concept definitions and against best practices for wording items. On this basis, aiming for a sweet spot between efficient and reliable measurement, we reduced the item set to four content-valid and balanced-keyed items per character strength. If necessary, we introduced alternate IPIP items from outside the IPIP-VIA item pool, mostly because some scales lacked suitable reversed items or needed improved wordings or better empirical psychometric properties (as determined in the next step). Third, after having translated or adapted the items to German in a thorough translation process, we tested the preliminary scales with participants from the UK and Germany. In these samples, we evaluated dimensionality, measurement model fit, and reliability. For a few scales, we developed alternative variants and compared the scale variants in two fresh samples from the UK and Germany. Fourth, we tested measurement invariance to ascertain cross-cultural applicability of the 24 final scales (henceforth IPIP-VIA-R). Finally, we investigated scale validity. We cast a nomological net with Big Five personality traits (Costa & McCrae, 1992; see Table 1.2a) and basic human values (Schwartz, 1992, 1994, 2006; see Table 1.2b). We also analyzed criterion validity for life satisfaction and health, next to gender differences in character strengths.

Table 1.2a

Facets Aligned with the Hierarchical Structure Yielding the Big Five Domains (sensu BFI-2)

Trait Label	Definition of Facet	Domain-Facet Abbreviation
Extraversion		
Sociability	Desire to socially approach and engage with others	E-s
Assertiveness	Willingness to express personal opinions and goals in social situations	E-a
Energy Level	Positive affect (especially positively aroused states: enthusiasm and excitement) and physical activity level	E-e
Agreeableness		
Compassion	Active emotional concern for others' well-being	A-c
Respectfulness	Treating others with regard for personal preferences and rights, inhibiting antagonistic/ aggressive impulses	A-r
Trust	Holding positive generalized beliefs about others	A-t
Conscientiousness		
Organization	Preference for order and structure	C-o
Productiveness	Work ethic and persistence while pursuing goals	C-p
Responsibility	Commitment to meeting duties and obligations	C-r
Negative Emotionality		
Anxiety	Tendency to experience anxiety and fear	N-a
Depression	Tendency toward depression and sadness	N-d
Emotional Volatility	Volatile mood swings	N-e
Open-Mindedness		
Intellectual Curiosity	Intellectual interests and enjoyment of thinking	O-i
Aesthetic Sensitivity	Broader alternative primarily defined by intellectual and artistic interests (+ other characteristics)	O-a
Creative Imagination	Creativity and originality	O-c

Note. Table based on Soto & John (2017a). Trait Label = Big Five domain (bold) or facet name (Negative Emotionality is historically known as Neuroticism or, reverse-coded Emotional Stability, and Open-Mindedness as Openness to [cognitive] experience); Domain = Big Five domain associated with a facet. The first facet for each domain is factor-pure, each next pair presents complementary facets (Open-Mindedness facets do not strictly conform to this rule). Underlining reflects letters used for acronyms for Big Five domain and facet names.

Table 1.2b

Definitions of Basic Human Values and Associated Value Clusters

Value Label	Definition of Value	Value Cluster
Power	Social status and prestige, control or dominance over people and resources (authority, social power, wealth, preserving my public image)	Self-Enhancement
Achievement	Personal success through demonstrating competence according to social standards (ambitious, successful, capable, influential)	Self-Enhancement
Hedonism	Pleasure or sensuous gratification for oneself (pleasure, enjoying life, self-indulgent)	Openness to Change
Stimulation	Excitement, novelty, and challenge in life (daring, a varied life, an exciting life)	Openness to Change
Self-Direction	Independent thought and action—choosing, creating, exploring (creativity, freedom, independent, choosing own goals, curious)	Openness to Change
Universalism	Understanding, appreciation, tolerance, and protection for the welfare of all people and for nature (equality, social justice, wisdom, broadminded, protecting the environment, unity with nature, a world of beauty)	Self-Transcendence
Benevolence	Preservation and enhancement of the welfare of people with whom one is in frequent personal contact (helpful, honest, forgiving, loyal, responsible)	Self-Transcendence
Tradition	Respect, commitment, and acceptance of the customs and ideas that traditional culture or religion provide (devout, respect for tradition, humble, moderate)	Conservation
Conformity	Restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms (self-discipline, politeness, honoring parents and elders, obedience)	Conservation
Security	Safety, harmony, and stability of society, of relationships, and of self (family security, national security, social order, clean, reciprocation of favors)	Conservation

Note. Table adapted from Schwartz & Boehnke (2004, Table 2). Value label = Basic human value sensu Schwartz; Value Cluster = Higher-order dimension pole. Underlining reflects acronyms for value cluster names.

We organize our report according to the research questions that we sought to answer: RQ1. What is the dimensionality of each character strength in the IPIP-VIA item pool, and which items represent each character strength's core? RQ2. Does the IPIP comprise a sufficient number of content-valid items per strength such that a selection of cross-culturally applicable items would form balanced-keyed and (essentially) unidimensional short scales? RQ3. Can IPIP-VIA-R measurement models based on a selection of four items per character strength achieve good fit? RQ4. Do brief balanced IPIP-VIA-R scales measure the character strengths reliably? RQ5. Are these IPIP-VIA-R short scales measurement invariant across two countries and languages? RQ6. Do the associations with relevant variables in the personality space support the construct validity of IPIP-VIA-R? RQ7. Do the IPIP-VIA-R scales possess criterion validity? Before we present the results for each RQ, we give a general overview over the scale development and validation process.

Overview Over the Scale Development and Validation Process

This section comprises two parts. The first part explains the source material, the shortening and refinement of IPIP-VIA strength measurement, and the translation/adaptation process. Further details on the item selection process and scale validation are available from OSF at https://osf.io/3mfyw/). The second part describes samples, sample size considerations, and study material.

Source Questionnaire for Self-Reported Character Strength Items: IPIP-VIA

The IPIP (Goldberg, 1999; Goldberg et al., 2006) is a large open-science repository of more than 3,200 personality questionnaire items (https://ipip.ori.org/HistoryOfTheIPIP.htm). IPIP scales are similar to widely used commercial personality inventories, yet they are placed in the public domain and hence free to use for all purposes without licensing charges or registration.

Peterson and Seligman's (2004) "Values in Action" (VIA) approach to surveying character features among the most widely used "human character" inventories. Around the year 2000, the IPIP-VIA scales comprised 240 items, with 10 items per character strength, except for Valor/Bravery (11 items) and Hope/Optimism (9 items). These items are largely identical to a preliminary version of the "VIA Inventory of Strengths" (VIA-IS) and were contributed by Peterson. Subsequently Goldberg (1999) reversed the keying of 3–4 items per scale and slightly adapted all items to fit the style and response format of the IPIP (see https://ipip.ori.org/Finding_Scales_to_Measure_Particular_Constructs.htm). In 2005, a slightly revised set of IPIP-VIA strength scales was developed (Diamond et al., 2010; Goldberg et al., 2006). With the aim to optimize Cronbach's Alpha (which, we argue, is a less-than-optimal criterion for scale development), 39 items were omitted and 12 new items added, leaving 213 recommended items at present. The resulting IPIP-VIA scales are (still) not balanced-keyed, and the proportion of reverse-keyed items varies across scales, making it impossible to adequately control for acquiescence. The IPIP webpage provides all 252 items (i.e., 240 original items including the 39 omitted items plus the 12 added items) that all served as source material in the present study (i.e., items omitted based on internal consistency considerations might still fulfill our item selection criteria of content validity and cross-cultural adaptability).

Revising and Refining IPIP-VIA Strength Measurement for Cross-Cultural Use

Reducing the Initial IPIP-VIA Item Set. Starting from the 252 IPIP-VIA items (https://ipip.ori.org/newVIAKey.htm), we set out to abbreviate the 24 strength scales by filtering out unsuitable items. Four independent expert raters inspected the items in order to (1) identify the four best (i.e., most content-valid and best-worded items) per strength while (2) avoiding redundancy within each scale and reducing overlap with other scales and (3) eliminating items that seemed prone to cultural bias or other cognitive errors in the response process. Another constraint was to ideally (4) retain the same number of positively and negatively worded items per scale. We found that the 252 IPIP-VIA items were insufficient to arrive at 24 content-valid and balanced-keyed strength scales (see RQs 1+2).

Adding Additional IPIP Candidate Items. To balance the scales with regard to content-valid and reverse-keyed items, we broadened the item pool by adding items from other IPIP sections (see SOM_IPIP-VIA-R_Method_Item-Source-History.xlsx). We screened 3,193 (non-VIA) items (variable sets termed A–Y),[1] of which 126 were potentially suitable for measuring VIA character strengths. Out of these, we chose seven items that matched the strength definitions, were linguistically sound, translatable, and free from cultural bias. We resolved lack of consensus in the item selection by comparing two questionnaire versions (Versions A and B; see SOM_IPIP-VIA-R_Method_Item-Selection.xlsx). In Version B, five items from five strength scales differed from Version A; additionally, five items from four strength scales obtained an improved translation (see RQ2).

Item Selection to Form Ultimate IPIP-VIA-R Questionnaire. We based the final decisions about which items from Versions A and B to include in the final 96-item IPIP-VIA-R on empirical results regarding dimensionality, model fit, and reliability. Among the decision criteria were factor loadings and goodness-of-model fit indices for acquiescence-controlled measurement models, closeness to unidimensionality, several reliability estimates, but also considerations of content validity and quality of translated items (see RQs 1–4).

TRAPD Approach: Item Translation and Adaptation

To ensure the cross-cultural applicability of the IPIP-VIA-R from the start, we jointly performed the final item selection in samples from two cultures and languages: English (US and UK) and German (Germany). Because the IPIP-VIA items were not available in German, we *adapted* the preselected English items for the German-speaking context. We followed Harkness's (2003) Translation, Review, Adjudication, Pretest, and Documentation (TRAPD) approach, which is the current gold standard translation approach. We also adhered to best-practice guidelines for translation and adaptation (Behr & Zabal, 2021). Three personality experts (native speakers of German and well-versed in English) independently prepared parallel translations. The translators plus two reviewers (native German survey methodologists with a strong command of English) agreed on the preferred item translations. An expert in cross-cultural studies and translation acted as an adjudicator during the review meeting and assessed the final versions vis-à-vis the source items. We documented all steps and decisions in the translation process (see SOM_IPIP-VIA-R_Method_TRAPD-Approach.xlsx and the accompanying documentation SOM_IPIP-VIA-R_Method_TRAPD-Approach.pdf).

During the adaptation process, we slightly improved translations compared to their English source versions where necessary, such as by eliminating minor inaccuracies in concept representation or by reducing overlap with neighboring scales through proper wording. We retrospectively coded any item alterations (from level 0 to level 4), with 0 = translating straightforwardly; 1 = addressing linguistic complexities; 2 = providing cultural adaptations (beyond linguistic issues); 3 = overcoming measurement-related issues (in the source item); 4 = aligning content more closely with the construct definition (potential misrepresentation of construct in the source). Table 1.3 provides an overview of the alterations (counting across the two variants we tested). These changes arguably brought the German adaptation closer to the construct definitions of the VIA character strengths, such that the German, rather than the English IPIP items, might be best suited to serve as a source for future translations.

For the response scales, we kept the IPIP's default of five response categories but slightly adjusted the English (and subsequently the German) response labels to strengthen the cross-cultural applicability of the response scale: 1 = *Does not apply at all* [*Trifft gar nicht zu*], 2 = *Applies only slightly* [*Trifft eher nicht zu*], 3 = *Applies moderately* [*Teils, teils*], 4 = *Applies mostly* [*Trifft eher zu*], 5 = *Applies completely* [*Trifft voll und ganz zu*].

Table 1.3

IPIP-VIA Item and Scale Adaptations During Revision and TRAPD Approach

IPIP-VIA Scale Adaptations	Occurrences at Item Level	Occurrences at Scale Level	% Scales Affected
Item Numbers			
Considered IPIP-VIA items	252	24	100.0
Dropped IPIP-VIA items before administration	156	24	100.0
Identified alternative (non-VIA) IPIP items	126	23	95.8
Tested (non-VIA) IPIP items	7	7	29.2
Preferred (non-VIA) IPIP items	6	6	25.0
Tested item variants (A vs. B) in UK & DE	5	5	20.8
Preferred item variants B over A in UK & DE	3	3	12.5
Tested additional translation variants (A vs. B) in DE	5	4	16.7
Preferred additional translation variants B over A in DE	5	4	16.7
Item Alterations			
Level 0: Translation straightforward	34	22	32.1
Level 1: Translation linguistically complex	38	22	35.8
Level 2: Adaptation culture-related	0	0	0.0
Level 3: Adaptation measurement-related	12	10	11.3
Level 4: Adaptation construct-related	22	12	20.8
Other item alterations	0	0	0.0
Response Scale Alterations			
Changed number of response categories	0	0	0.0
Changed scale anchors (response labels)	96	24	100.0
Other Adaptations	0	0	0.0

Note: Percentages do not sum to 100% due to multiple adaptations of the scales.

Samples and Data Quality

We used five samples from three nations during three stages of the item selection and scale validation process: In the first stage, we used an extant US community-dwelling sample (98.4% White non-Hispanic or Euro-American, reported as “Caucasian”, which reflects the ethnic composition of homeowners around 1993, that is, at the time of creating the panel). In the second stage, we collected two samples of online panelists who were recruited simultaneously in the UK and Germany via commercial online access panels. In the third stage, we used two additional independent online-samples from the UK and Germany. We obtained respondents’ informed consent by agreement to the panel provider’s study invitation (or by signature to the original data collectors in the United States). We chose sample sizes of around 500 to ensure sufficient (~90%) statistical power for testing the intended measurement models (see SOM_IPIP-VIA-R_RQ3_Sample-Size_Power-Analysis.pdf). Online samples were anonymously collected in two waves without tracking participants’ ethnicity. (Mirroring their society, most of them were likely White, too, though specific minority proportions remain unknown.) Table 1.4 shows the sample compositions.

Sample 1. The Oregon Research Institute (ORI) recruited a sample of approximately 750 US homeowners in the Eugene–Springfield (Oregon) community who have completed IPIP items and a large set of other psychological measures. These measures were collected in multiple survey waves that encompass several years. (Not all Sample 1 respondents completed all sections though). We used extant data from this “Eugene-Springfield Community Sample” (ESCS; Goldberg & Saucier, 2016) for initial scale inspection and shortening. Participants had worked on a paper-pencil interview on “Perceptions of Personal Qualities” (PPQ) in 2004 in exchange for a usual remuneration (roughly $15 at the time). Data from 713 participants on 252 IPIP-VIA items were available (711 complete cases for factor analyses, because two participants had accidentally skipped 29 out of 342 PPQ-items items *en bloc*, and one of them produced another isolated missing value). A varying case number (with up to 126 missing cases) resulted when analyzing the non-VIA IPIP items added retroactively to the present-day IPIP-VIA item set.

Samples 2 and 3. In January 2018, a total of 1,021 native speakers from the UK and Germany completed the online survey, which allowed a glance at the item selection including non-VIA items that were taken, for the first time, by all study participants and at the same time. The samples were drawn by a commercial online access panel provider (respondi AG) in exchange for a small monetary compensation (roughly 2€). Sample 3 matched the register-based German census proportions cross-classifying gender, three cohorts (age 18-29, 30-49, 50-69) and three educational levels attained (basic, middle, high education; Gauckler & Körner, 2011). For comparable subsamples, respondi drew a parallel (i.e., matched quota) sample in the UK (sample 2). Completion rates were lower in the UK (27.11%) than in Germany (49.85%), hinting at higher motivation and potentially higher data quality among the German panelists. After filtering out respondents with poor quality data, the analytical samples comprised 921 respondents (455 from the UK, 466 from Germany).

Samples 4 and 5. For finalizing the item selection and validating measurement models across countries, we recruited two additional samples (total N = 1,040) via respondi in October 2018, with none of the respondents having participated in the earlier wave. Again, there were differences in completion rates between the UK (33.14%; sample 4) and Germany (53.68%; sample 5). Filtering out low quality data, the analytical samples comprised 950 respondents (474 from the UK, 476 from Germany).

Quality Control (Filtering). For the secondary data analysis of sample 1 (paper-and-pencil questionnaire), we kept all responses of the existing data set. For samples 2 through 5 (online surveys), we followed suggestions to filter out careless responders in line with Meade and Craig (2012). In the absence of scientific consensus for a standard procedure (Leiner, 2019), we excluded participants who did not pass *minimum* quality thresholds but not more than 10% of each sample, by combining three non-reactive indicators[2]: (1) To filter out low internal motivation (rushing/patterning), we kept only those panelists who spent more than an average second per relevant personality item according to server-side recorded time-stamps for each questionnaire page. (2) We computed each respondent's ipsative variance across all their ratings on strength and other personality items. We dropped 5% of participants in the lower tail of each country's score distribution, as low ipsative variance indicates straight-lining or non-differentiation. (3) Across the same items, we computed a robust multivariate Mahalanobis distance, *D*, that can handle missing data (Béguin & Hulliger, 2004). *D* quantifies the distance of an individual's response vector from the average response vector and indicates deviant response patterns. A high distance indicates highly unusual response patterns, such as due to inattentive responding or faking. We removed per country 2.5% of participants with *D*-values exceeding the 97.5%-quantile. For details on the quality-enhancing procedure, see SOM_IPIP-VIA-R_Method_Data-Quality.pdf.

Table 1.4

Socio-demographic Descriptives of Samples 1-5

	Sample 1	Samples 2	Sample 3	Sample 4	Sample 5
	ESCS	UK	DE	UK	DE
Gender					
Female	403	259	256	257	262
Male	304	249	257	265	256
Marginals	713	508	513	522	518
Total Survey Completers	713	1021		1040	
Analytical Sample	713	455	466	474	476
Analytical Total Survey	713	921		950	
Age					
Range	18–83	18–69	18–69	18–69	18–69
M	49.90	44.30	43.52	43.48	43.65
SD	12.02	14.60	14.31	14.34	15.14

Note. Samples 1–5 = available survey completers only. ESCS = Eugene-Springfield Community Sample (six gender values missing); UK = United Kingdom; DE = Germany; Analytical Sample = after filtering out participants for poor data quality (see text). Samples 2 and 3 (IPIP-VIA-R Version A) comprised 7 missing data points, Samples 4 and 5 (IPIP-VIA-R Version B) comprised 2 missing data points.

Material

All surveys involved multiple item-batteries, questions, and constructs unrelated to the current study, which we will not cover here. For details on specific validation criteria (e.g., Big Five and Schwartz' basic human values), we refer the reader to RQ6 and RQ7.

IPIP-VIA Questionnaire for Sample 1. In Fall 2004, as part of the PPQ-study with 342 personality items, the ESCS-participants were also administered IPIP-VIA items.[3] We re-analyzed all 252 IPIP-VIA items available on the IPIP webpage. Recall that the IPIP-VIA webpage presently recommends using a subset of 213 IPIP-VIA items. However, as the remaining 39 items were omitted to enhance the scales' internal consistency, they might still fulfill our item selection criteria (i.e., content-valid, cross-cultural adaptable).

IPIP-VIA-R (Version A) for Samples 2 and 3. Participants worked on 96 items for IPIP-VIA-R short scales (Version A) that we distributed across ten pages of a longer 200-item online survey. In both countries, the IPIP-VIA-R section took about 9 min. to complete.

IPIP-VIA-R (Version B + A) for Samples 4 and 5. Rater disagreement about item selection and our initial empirical evaluation of Version A necessitated a crosscheck of scales with five alternative items and another five new translations of original items (Version B). Furthermore, because the IPIP-VIA-R items were presented towards the end of a longer survey in Samples 2 and 3, fatigue might have impaired data quality. Therefore, in Samples 4 and 5, that we used as the ultimate basis for the psychometric validation of IPIP-VIA-R, we presented Version B of IPIP-VIA-R as the first substantive part following a short block of generic questions (e.g., socio-demographic questions). To allow for a comparison between Versions A and B, we also included the items from Version A immediately after the 96 Version B items (UK: five original items; Germany: five original items plus five original translations). The B-versions replaced their precursors at the respective questionnaire positions of Version A. Working on 101 (106 German) items lasted 10 min. (12 min.) with an average completion time of about 20 min. for the survey.

Data Preparation and Statistical Analysis. We performed the preparatory steps and analyses with open- and closed-source software. We ran descriptive analyses and unconstrained factor analyses in SPSS (version 24.0) and FACTOR (version 10.8.04; Lorenzo-Seva & Ferrando, 2006, 2013), confirmatory factor analyses in Mplus (version 7.31; L. K. Muthén & Muthén, 1998–2012) with the *R*-package *MplusAutomation* (version 0.7-3; Hallquist & Wiley, 2018) as well as with the *R*-package *lavaan* (version 0.6-5; Rosseel, 2012) in *R* 3.4.4 (R Core Team, 2018). We gained additional information from the *R*-package *psych* (version 1.8.12; Revelle, 2019b). Validity correlations were computed in Stata (version 15) and *R*.

Research Question 1:
What Is the Dimensionality of Each Character Strength in the IPIP-VIA Item Pool, and Which Items Represent Each Character Strength's Core?

Items should reflect the relevant target dimension (i.e., the common core of the construct). Yet, when factor-analyzing all VIA-IS strength items jointly (McGrath, 2014), many items show only small loadings on the assigned strength factor ($\lambda < .40$), indicating rather loose item-factor relationships (Hair et al., 2006). Given the statistical evidence and conceptual criticism of other VIA character strength inventories (see also above), we tested whether—as we assumed—the same issues also plague the 252 IPIP-VIA items. Expecting confirming evidence in this regard, we sought to statistically identify suitable candidate items at the same time. We explored dimensionality and factor loadings for the itemset of each IPIP-VIA strength through exploratory factor analyses (EFA) separately for each strength (Ferrando & Lorenzo-Seva, 2018; Saucier & Goldberg, 1996). Later, we tested measurement models for short scales via confirmatory factor analyses (CFA) with fresh samples (i.e., Samples 2–5).

Factor analysis casts light on the dimensionality of an item set, and unidimensionality is a highly desirable measurement property particularly for short scales (Bond & Fox, 2015;

Hattie, 1985). Lack of unidimensionality complicates or even invalidates the common practice of summating across ratings and interpreting the test scores as if they reflect a single construct. Measures of internal consistency such as Cronbach's Alpha also require (but do not test) unidimensionality and are biased if unidimensionality does not hold (McNeish, 2018; Raykov, 2001). Strict unidimensionality requires the absence of secondary factors (in EFA) or correlated residuals (in CFA), which may be rooted in identical keywords or grammatical constructions. However, personality items are often factorially complex and rarely behave as markers of a single factor (Cattell, 1986), especially if items and scales are not thoroughly constructed. *Essential* unidimensionality is then a more realistic goal (Stout, 1987). It entails that the common factor explains most variability in scores, while additional factors (e.g., method factors) explain only small variance portions (or item-pair covariances) that do not harm measurement, at least when properly specified in a measurement model. Attaining (essential) unidimensionality often requires item selection and/or model re-specification (Brown, 2015; Gerbing & Anderson, 1988; Hopwood & Donnellan, 2010).

To gauge the extent of (deviation from) unidimensionality in the IPIP-VIA—and to identify positively and negatively keyed candidate items for balancing the scales—we analyzed each of the 24 character strengths as measured with all available 252 IPIP-VIA items. For this purpose, we ran a secondary data analysis on Sample 1.

Previous issues about dimensionality demonstrated almost exclusively for VIA-IS character strength measures are likely to re-emerge within IPIP-VIA item sets. The degree to which the item set of each character strength deviates from unidimensionality likely coincides with the extent of (1) poor wording or construct representation of items, (2) present item clusters or "sub-facets", and (3) method variance due to acquiescent responding and presence of negatively keyed items.

Method

In a first step, we examined descriptive statistics for each item (details in SOM_IPIP-VIA-R_RQ1_Descriptives.pdf). The skew of the variables rarely exceeded an absolute value of +2, while absolute excess kurtosis values of +3 occurred (where zero can be expected for normal distributions). Yet, Bartlett's tests of sphericity and sampling adequacy according to the Kaiser–Meyer–Olkin criterion (KMO = .78–.92; Dziuban & Shirkey, 1974; Kaiser & Rice, 1974) supported the suitability of the 24 scales for factor analysis (see Table 1.5).

We judged unidimensionality by (1) parallel analysis (Horn, 1965), which compares Eigenvalues obtained from principal axis factoring (PAF) and principal component analysis (PCA) against the 95th percentiles of Eigenvalues resulting from 500 draws of identically structured random data; (2) the minimum average partial (MAP) test (Velicer et al., 2000); (3) factor loadings on the common single factor, indicating the quality of unidimensional measurement; (4) item communality when extracting multiple factors (with initial Eigenvalues > 1) after applying orthogonal Varimax rotation (the most parsimonious explanation of item correlations by independent factors) as well as the accompanying proportions of variance captured by the *k* extracted factors (indicating of the size of potentially relevant secondary factors).[4]

Table 1.5

Cutoff Criteria for Factor Analytical Procedures in Scale Development (Subject to Professional Judgment)

Index (Shorthand)	Index Definition and Relevance	General Rules for Acceptability	Cutoff Applied to Short Scales (IPIP-VIA-R)
Suitability of Factor Analysis (EFA, PCA)			
KMO	Kaiser-Meyer-Olkin criterion of overall sampling adequacy: suitability of a data matrix prior to any factoring procedures	KMO < .50 ("unacceptable") – Kaiser & Rice (1974) KMO ≥ .60 – Dziuban & Shirkey (1974) KMO = .60 ("mediocre") – Lorenzo-Seva & Ferrando (2013) KMO ≥ .60 – Tabachnick & Fidell (2013)	KMO ≥ .60
Unidimensionality			
ECV	Explained Common Variance: variance explained by first (common) Minimum Rank Factor Analysis factor divided by the total common variance, i.e., the dominance of the first factor over other factors (or relative general-factor strength index)	ECV ≥ .70–.85 ("essentially unidimensional") – Ferrando & Lorenzo-Seva (2018), following Green et al. (1984) and Rodriguez et al. (2016)	ECV ≥ .60
MIREAL	Mean Item Residual Absolute Loading: average of item loadings on secondary factor	MIREAL ≤ .30 ("rough initial reference" for essential unidimensionality) – Ferrando & Lorenzo-Seva (2018), following Grice (2001) Strength of loading indicating a factor: .32 ("poor"), .45 ("fair"), .55 ("good"), .63 ("very good") or .71 ("excellent") – Tabachnick & Fidell (2013), following Comrey & Lee (2013)	MIREAL ≤ .40

(continues)

Index (Shorthand)	Index Definition and Relevance	General Rules for Acceptability	Cutoff Applied to Short Scales (IPIP-VIA-R)
Factor Reliability & Construct Replicability			
FDI	Factor Determinacy Index: correlation between factor scores and intended factor (congeneric latent variable)	FDI ≥ .80 (“adequate”) – Gorsuch (1983–2015) FDI ≥ .90 (“recommended”) – Brown (2015), Beauducel (2011)	FDI ≥ .80
H / gHI	H-Index: proportion of variance explained by congeneric latent variable indicators (Hancock & Mueller, 2001), also termed maximal reliability (internal consistency for an optimally weighted scaling procedure) gHI is the generalized H-Index applicable to categorical data, also in the presence of secondary (even oblique) factors; it may refer to *latent* response variables or *observed* variables (Ferrando & Lorenzo-Seva, 2018); the expected gHI values are lower for observed than for latent response variables (the latter are merely hypothetical)	H ≥ .70 (“reasonable”) – Hancock & Mueller (2001) H ≥ .80 (“well-defined latent variable”) – Rodriguez et al. (2016) gHI_{latent} ≥ .80 (“reasonable”) – Ferrando & Lorenzo-Seva (2018)	$gHI_{observed}$ ≥ .65
Scale Reliability			
AVE	Average Variance Extracted: average ratio of item variance explained by the respective construct (and identified secondary sources of variability), as a conservative approach to the overall validity of a measurement model (also referred to as “convergent validity” from the perspective of the indicators)	AVE ≥ .50 – Fornell & Larcker (1981) AVE < .50 (“acceptable”) if composite reliability is high enough: - CR ≥ .60 – Fornell & Larcker (1981) - CR ≥ .70 – Hair et al. (2006)	AVE ≤ .50 if: CR (= ω) ≥ .60
ω (or CR)	Omega: proportion of variance in the unit-weighted scale score attributable to the common factor (but not group-factors or item-specific factors) hence a measure of general factor saturation (McDonald, 1999), also termed Composite Reliability (Fornell & Larcker, 1981) or Raykov’s (1997) rho; unlike Cronbach’s α, ω does not depend on the assumption of essential tau-equivalence and is unbiased if multidimensional items are properly specified	Reliability norms for ω ≈ α α ≥ .90 (“excellent”), .80 (“good”), .70 (“acceptable”), .60 (“questionable”), or .50 (“poor”) – Nunnally & Bernstein (1994) α ≈ .70 – Schmitt (1996) denied that α must always exceed .70	ω ≥ .70
r_{tt}	Test-retest reliability: rank-order consistency (not: repeatability or agreement of successive measurements)	r_{tt}: cutoff is “circumstantial” – Crocker & Algina (1986) r_{tt} ≥ .80 (assuming character strengths are stable traits)	r_{tt} > .70

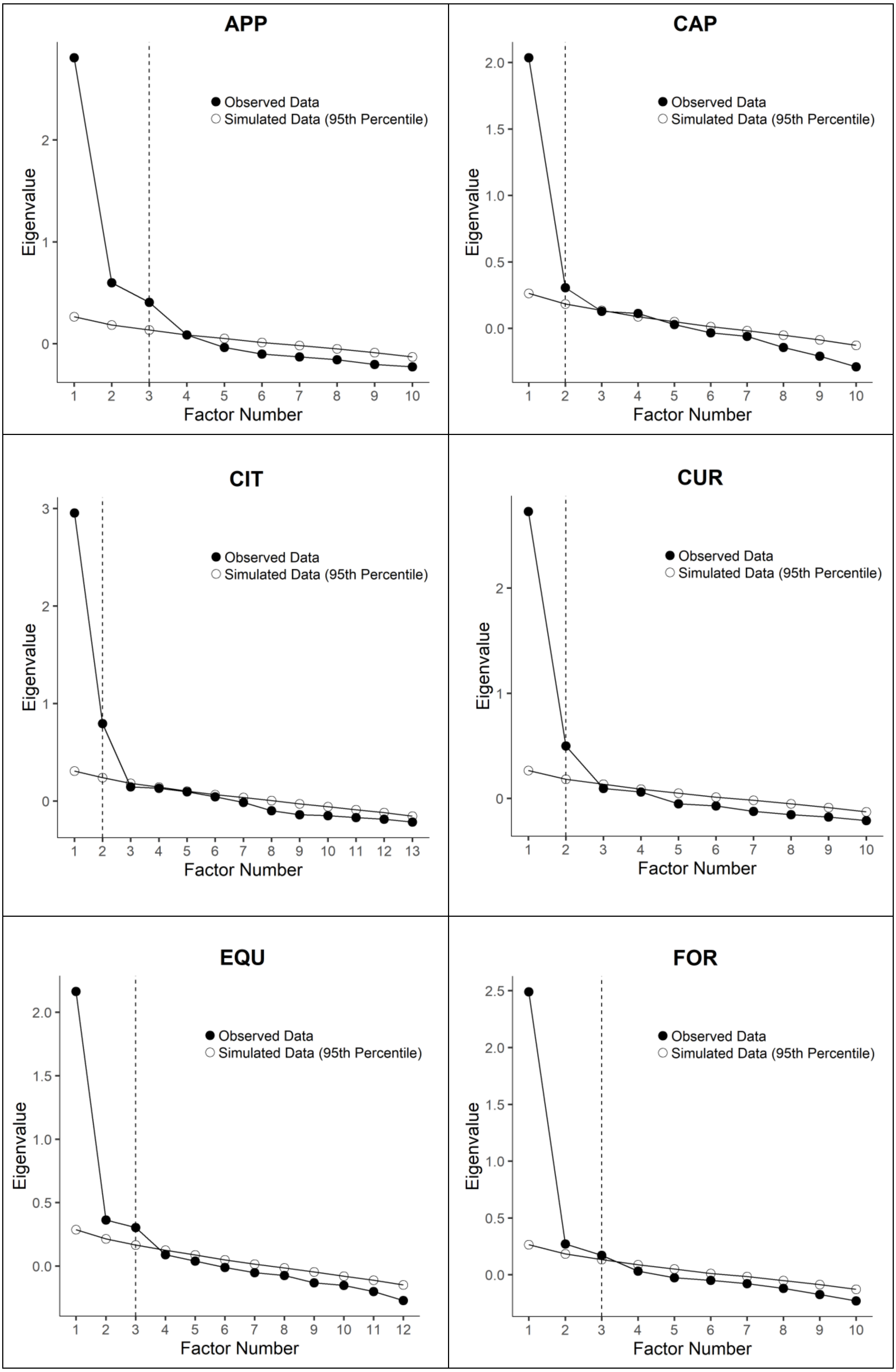
(continues)

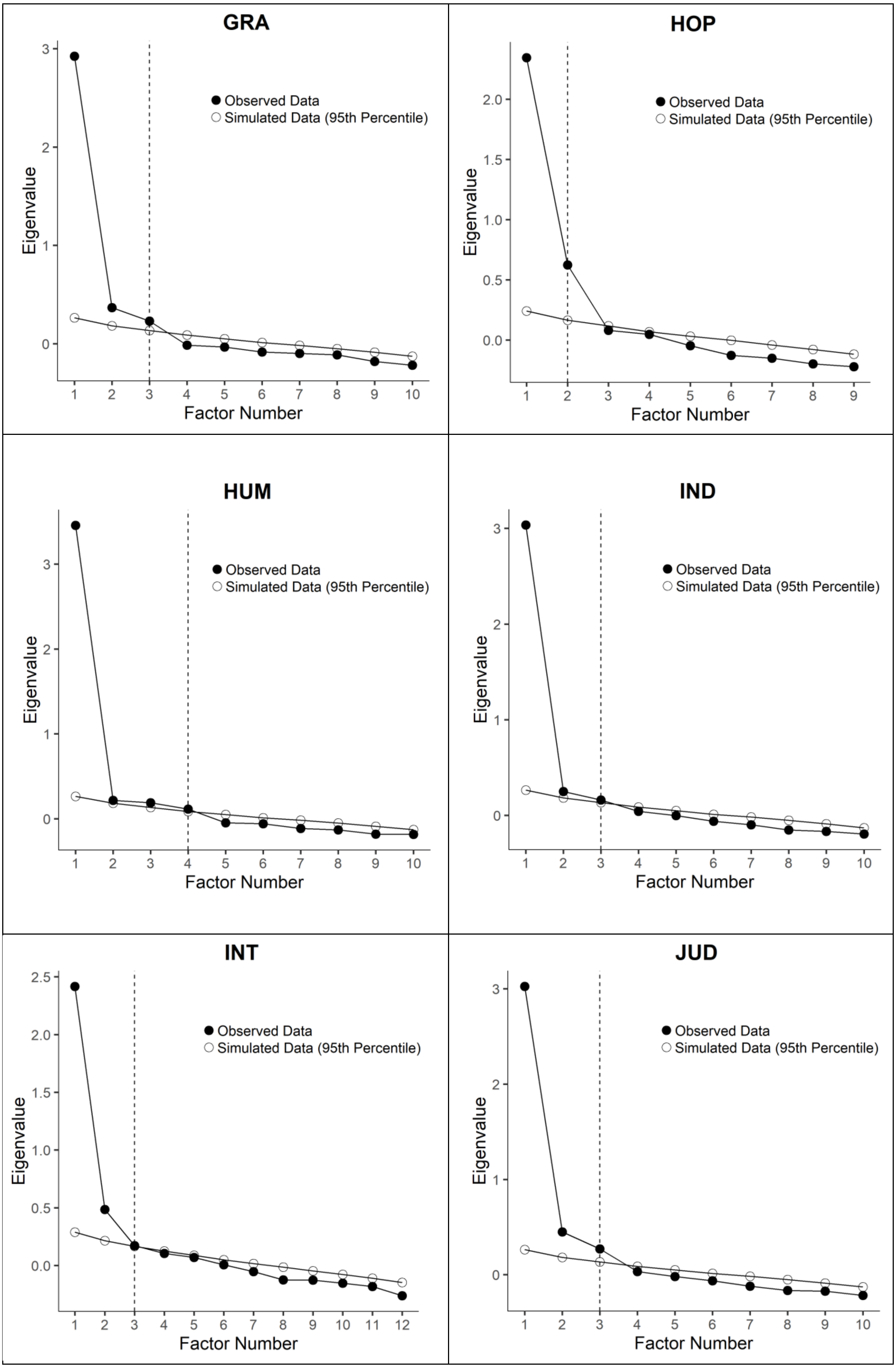
Index (Shorthand)	Index Definition and Relevance	General Rules for Acceptability	Cutoff Applied to Short Scales (IPIP-VIA-R)
<u>Model Fit</u>			
χ^2_{SB}	Chi-square statistic: Satorra-Bentler scaled for using robust ML (MLR), with *df* being the expected value	nonsignificant	nonsignificant
χ^2_{SB} / *df*	Normed chi-square: Ratio of chi-square to degrees of freedom	χ^2_{SB} / *df* ≤ 2–3 – Kline (2011) χ^2_{SB} / *df* ≤ 5 – Schumacker & Lomax (2004)	χ^2_{SB} / *df* ≤ 10
CFI	Comparative Fit Index: incremental measure of fit in comparison to null model (Bentler, 1990)	CFI ≥ .95 – Hu & Bentler (1999) CFI < .90 ("poor fit") – Bentler & Bonett (1980)	CFI ≥ .90
RMSEA	Root Mean Square Error of Approximation: measure of approximate fit based on the discrepancy per degree of freedom (Steiger, 1990)	RMSEA ≤ .06 – Hu & Bentler (1999) RMSEA ≤ .10 ("mediocre") – Browne & Cudeck (1993)	RMSEA ≤ .10
SRMR	Standardized Root Mean Square Residual: discrepancy between the standardized sample covariance matrix and the model covariance matrix (Hu & Bentler, 1999); alternatively: discrepancy between the observed and estimated correlations, standardized means, and variances (Asparouhov & Muthén, 2018)	SRMR ≤ .05 (Byrne, 1998) SRMR ≤ .08 (Hu & Bentler, 1999)	SRMR ≤. 08
BIC	Bayesian Information Criterion, allows (nested and nonnested) model comparisons	the smaller the better	smaller

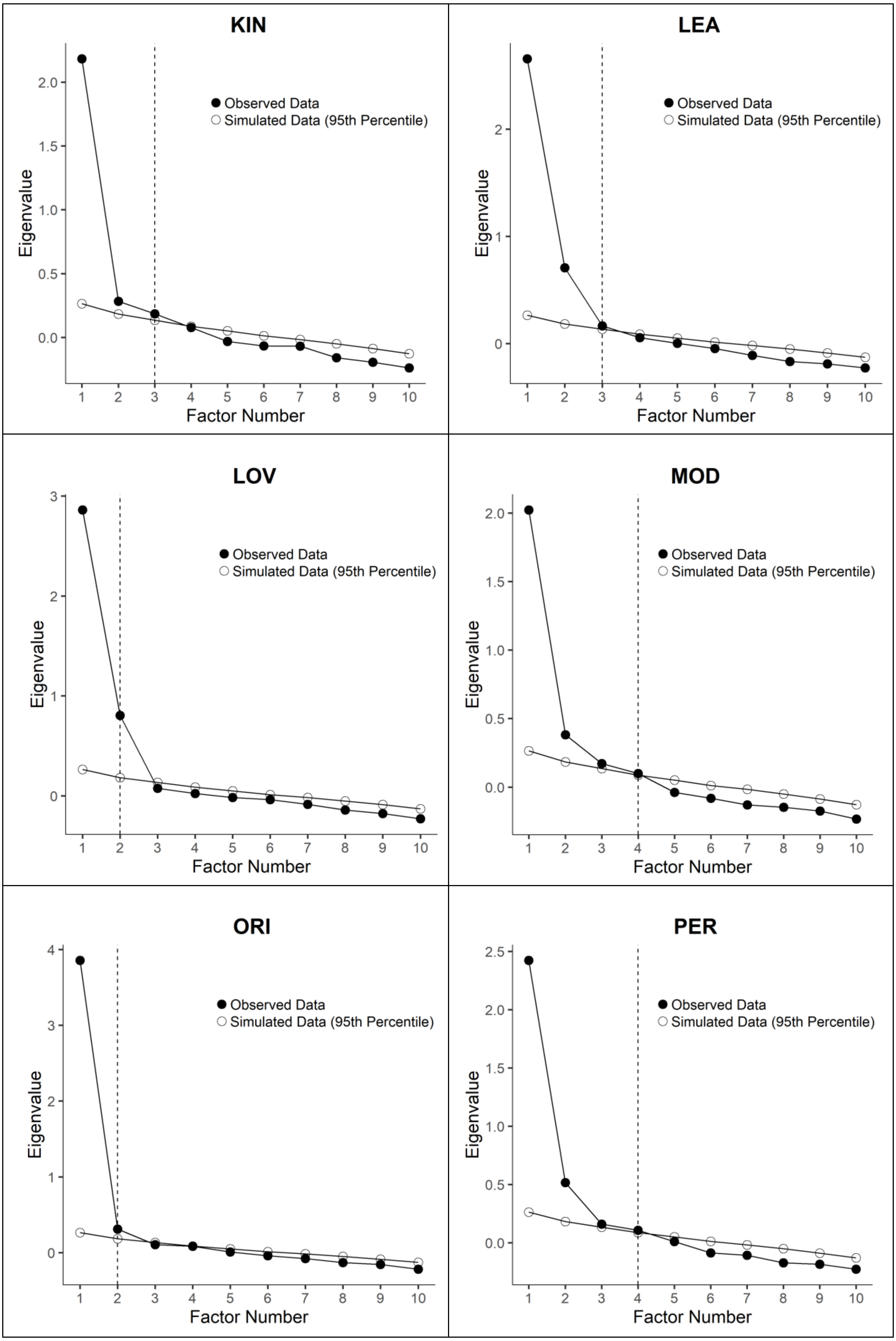
Note. Instead of using *Mplus*'s CFI and RMSEA (or scaled versions) for evaluating the initial measurement models, we relied on the nonnormality-robust, sample-appropriate indices that reflect true population characteristics: $\text{CFI}_{\text{robust}}$ and $\text{RMSEA}_{\text{robust}}$ (Brosseau-Liard et al., 2012; Brosseau-Liard & Savalei, 2014), which at the time of the study, were available only from *R/lavaan*. $\text{RMSEA}_{\text{robust}}$ will usually be higher than $\text{RMSEA}_{\text{scaled}}$ or RMSEA, whereas $\text{CFI}_{\text{robust}}$ can be higher or lower than $\text{CFI}_{\text{scaled}}$ or CFI (Savalei, 2018). For invariance testing though, we used *Mplus*'s $\text{CFI}_{\text{scaled}}$ and $\text{RMSEA}_{\text{scaled}}$ for, because the prominent cut-off heuristics for inspecting model *differences* to determine nonequivalence were derived from simulations using the seminal fit indices and because we considered information criteria (BIC) to be more relevant for judging the acceptability of parsimonious measurement invariance models anyway.

Results

Visual inspection of Eigenvalues (PAF scree plots) revealed that most item sets had a strong first factor, after which Eigenvalues leveled off (see Figure 1.1). The first factors explained between 26% (Self-Regulation) and 56% (Spirituality/Religiousness) of the variance in each strength (see SOM_IPIP-VIA-R_RQ1_Dimensionality.pdf). At the same time, parallel analysis suggested that more than three dimensions were present in 18 of the item sets, and a secondary dimension was likely present in the remaining 6 item sets. For instance, whereas a single dimension *seemed* sufficient for Originality, Spirituality, Humor, and Prudence according to the MAP-test, PAF parallel analysis suggested two, three, four, and five common factors to underlie the respective strengths. Although MAP often suggested fewer dimensions than parallel analysis (see Table 1.6), it should be noted that underfactoring (specifying too few factors) is considered a more severe error than overfactoring (specifying too many factors; Fabrigar et al., 1999). Also note that acquiescence can bias the number of factors to retain upward in classical random parallel analysis, but downward in the MAP-test and permutation parallel analysis (Valentini, 2017). The variance attributed to a second or third factor was often non-negligible (e.g., after rotation two factors equally contributed 15% of common variance to Forgiveness; Appreciation of Beauty and Excellence comprised three factors with common variances of 16%, 15%, and 11%).


APP
Observed Data
Simulated Data (95th Percentile)
Eigenvalue
Factor Number
CAP
Observed Data
Simulated Data (95th Percentile)
Eigenvalue
Factor Number
CIT
Observed Data
Simulated Data (95th Percentile)
Eigenvalue
Factor Number
CUR
Observed Data
Simulated Data (95th Percentile)
Eigenvalue
Factor Number
EQU
Observed Data
Simulated Data (95th Percentile)
Eigenvalue
Factor Number
FOR
Observed Data
Simulated Data (95th Percentile)
Eigenvalue
Factor Number


GRA
HOP
HUM
IND
INT
JUD
Observed Data
Simulated Data (95th Percentile)
Eigenvalue
Factor Number


KIN
Eigenvalue
2.0
1.5
1.0
0.5
0.0
Observed Data
Simulated Data (95th Percentile)
1 2 3 4 5 6 7 8 9 10
Factor Number
LEA
Eigenvalue
2
1
0
Observed Data
Simulated Data (95th Percentile)
1 2 3 4 5 6 7 8 9 10
Factor Number
LOV
Eigenvalue
3
2
1
0
Observed Data
Simulated Data (95th Percentile)
1 2 3 4 5 6 7 8 9 10
Factor Number
MOD
Eigenvalue
2.0
1.5
1.0
0.5
0.0
Observed Data
Simulated Data (95th Percentile)
1 2 3 4 5 6 7 8 9 10
Factor Number
ORI
Eigenvalue
4
3
2
1
0
Observed Data
Simulated Data (95th Percentile)
1 2 3 4 5 6 7 8 9 10
Factor Number
PER
Eigenvalue
2.5
2.0
1.5
1.0
0.5
0.0
Observed Data
Simulated Data (95th Percentile)
1 2 3 4 5 6 7 8 9 10
Factor Number

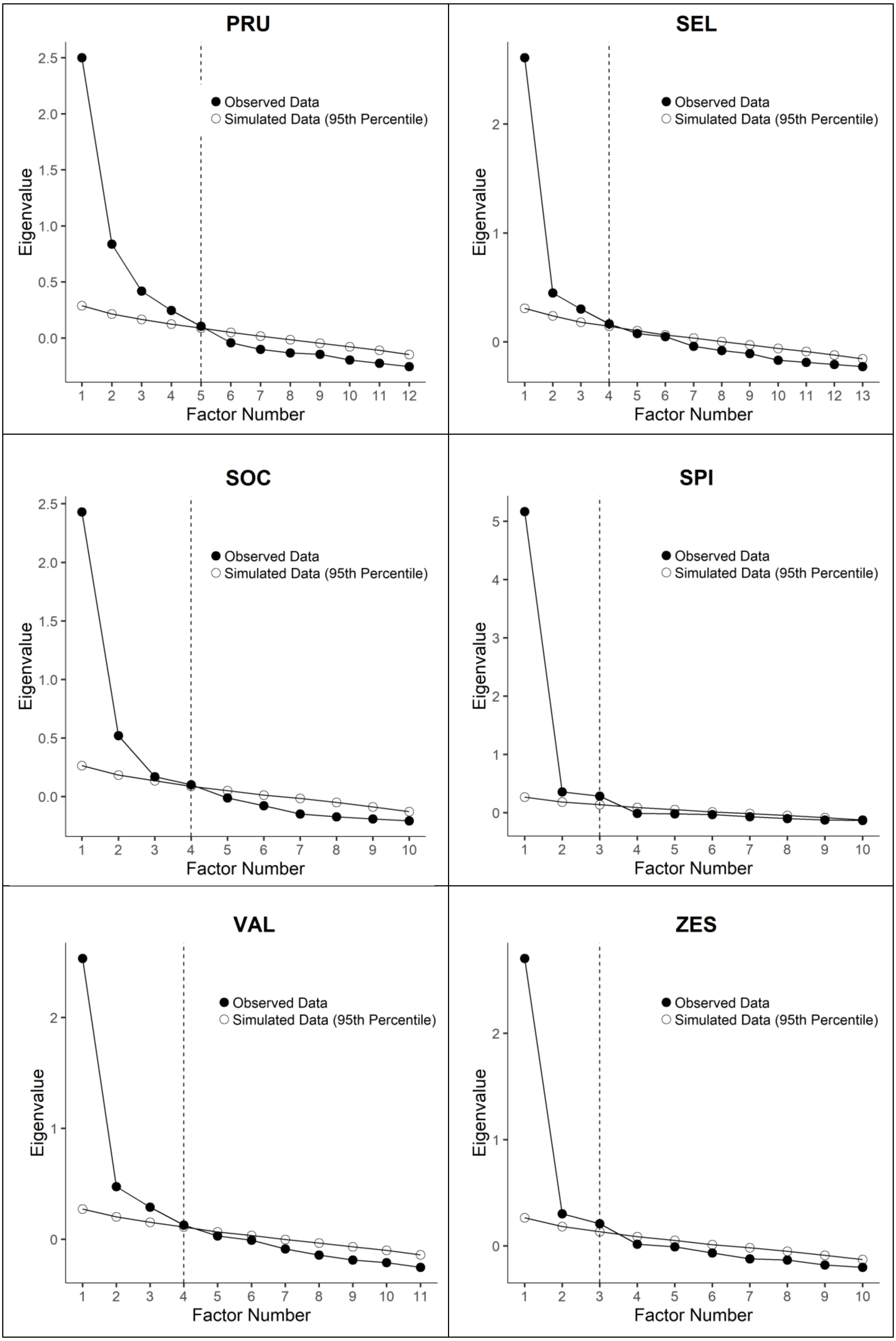


Figure 1.1. Scree-plots from 24 principal axis factor analyses of IPIP-VIA strengths with number of factors as indicated for each by parallel analyses (Sample 1: ESCS).

Table 1.6

IPIP-VIA Dimensionality in Full (Sample 1) and Abbreviated IPIP-VIA Strength Scales (Variant A for Samples 2 and 3; Variants A/B for Samples 4 and 5)

Scale	Full Scale: Sample 1			Abbreviated Scale Variant A: Samples 2 and 3						Abbreviated Scale Variant A/B: Samples 4 and 5					
	MAP-Test	PA-PCA	PA-PAF	PA-MRFA		ECV		MIREAL		PA-MRFA		ECV		MIREAL	
	ESCS	ESCS	ESCS	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE
APP	1	3	3	0	0	.53	.81	.48	.19	0/0	0/1	.54/.63	.83/.74	.46/.42	.18/.29
CAP	1	1	2	0	1	.71	.81	.32	.24	1	1	.73	.82	.33	.27
CIT	1	2	2	1	0	.70	.65	.41	.33	2	0/1	.70	.63/.68	.38	.41/.34
CUR	1	2	2	1	1	.74	.76	.36	.34	1	1	.73	.76	.34	.33
EQU	1	1	3	1	0	.74	.82	.35	.20	1	1/1	.77	.79/.76	.34	.26/.30
FOR	1	1	3	1	1	.69	.77	.37	.30	0	1/1	.67	.71/.84	.37	.37/.24
GRA	1	1	3	0	1	.68	.73	.40	.32	1	1	.69	.79	.41	.26
HOP	1	2	2	1	0	.69	.67	.41	.38	0	1	.60	.69	.48	.40
HUM	1	1	4	1	1	.72	.82	.34	.31	0	1	.68	.81	.40	.29
IND	1	1	3	1	1	.78	.85	.32	.28	1	1	.76	.94	.35	.15
INT	1	2	3	1	1	.77	.81	.34	.27	1	1	.73	.88	.39	.23
JUD	1	2	3	1	0	.70	.67	.35	.34	0	1	.66	.72	.41	.31
KIN	1	1	3	1	1	.67	.83	.37	.25	2	1	.69	.70	.39	.33
LEA	1	2	3	1	1	.75	.79	.35	.30	0	1	.63	.75	.40	.33
LOV	2	2	2	0	1	.63	.75	.34	.31	0/1	1/1	.63/.78	.71/.89	.34/.31	.33/.17
MOD	1	2	4	0	1	.64	.81	.37	.24	0	0	.62	.70	.43	.31
ORI	1	1	2	1	1	.75	.86	.37	.23	1	1	.69	.81	.42	.28
PER	1	2	4	1	1	.70	.72	.35	.34	0	1	.67	.70	.40	.35
PRU	1	3	5	0	1	.61	.72	.42	.35	0/1	1/0	.59/.73	.73/.71	.41/.30	.32/.27
SEL	1	2	4	0	1	.60	.71	.38	.26	0	0/0	.59	.61/.65	.40	.34/.32
SOC	1	2	4	0	1	.66	.76	.40	.30	1	0	.67	.69	.43	.36
SPI	1	1	3	0	0	.60	.67	.42	.44	0/0	1/0	.66/.64	.73/.64	.38/.39	.37/.35
VAL	1	2	4	0	1	.66	.90	.38	.18	0	1	.66	.86	.40	.20
ZES	1	1	3	1	1	.64	.77	.37	.33	0/1	1/1	.68/.67	.77/.75	.40/.39	.33/.35

Note: MAP-Test = Velicer's Minimum Average Partial Test, PA-PCA: Parallel Analysis based on Principal Component Analysis; PA-PAF: Parallel Analysis based on Principal Axis Factor Analysis; PA-MRFA: Parallel Analysis based on Minimum Rank Factor Analysis; ECV: Common Variance explained by first factor; MIREAL: Mean Item Residual Average Loading.

Extracting a single-factor per strength in EFA, most loadings were of modest size (see SOM_IPIP-VIA-R_RQ1_Dimensionality.pdf). Twenty items loaded lower than $\lambda < .25$, and some loadings were low enough to disqualify the attendant items as indicators of the common strength factor (e.g., “Have never given bad advice to a friend” obtained only a standardized loading of $\lambda = .10$ on the Perspective factor). Even when disregarding the 39 out of 252 IPIP-VIA items not recommended on the IPIP webpage (to improve Cronbach’s alpha), 28 of the 213 (i.e., more than 10%) currently used IPIP-VIA items had loadings below $\lambda = .40$, indicating rather loose item-factor relationships (Hair et al., 2006). Our findings coincide with McGrath’s (2014), who showed that also in the VIA-IS, a non-negligible number of items loaded < .40 on the strengths they were supposed to measure in a joint item-factor analysis combining all the VIA-IS strength items. The loadings already imply some heterogeneity in the item sets and somewhat weak item-total correlations for several items.

The few reverse-keyed items provided by IPIP-VIA would already make it challenging to arrive at high-quality balanced-keyed strength scales. However, not only that item sets comprised few reversed items *per se*, but the items also often insufficiently captured the strength’s core, as reflected in poor standardized loadings. For example, of three negative Forgiveness items “[I] hold grudges” works well ($\lambda = -.71$), but the loadings for “[I] am unwilling to accept apologies” ($\lambda = -.25$) and “[I] do not give anyone a second chance to hurt me” ($\lambda = -.31$) lay a shadow over our ability to arrive at balanced-keyed scales, even when opting for merely *four* IPIP-VIA items per strength.

Extracting between two and five factors with EFA (depending on the Eigenvalues for each strength) made it clear why this was the case. Several items had low communalities, suggesting that these items fall outside the trait space of their targeted strength. For instance, researchers are discouraged on the IPIP-VIA webpage to use the item “[I] could never stop loving my family and close friends, no matter what they did” for measuring Capacity for Love, a recommendation we confirmed through a disappointingly low communality (h^2 =

.06). However, similar weaknesses existed in items that *are* recommended for strength measurement with IPIP-VIA, such as “[I] admit when I am wrong” (Equity, h^2 = .12) and “[I] am not confident that my way of doing things will work out for the best” (Hope, h^2 = .15), though other recommended items fared comparatively better (h^2 = .25–.56). At the same time, some items are not recommended for strength measurement with IPIP-VIA despite reaching similar and reasonable levels of communality (e.g., [I] am rarely aware of the natural beauty in the environment”, Appreciation of Beauty and Excellence, h^2 = .25).[5]

Discussion

To conclude, many IPIP-VIA items currently recommended on the IPIP website appear problematic from a purely statistical (factor-analytic) standpoint already. The assumption of unidimensionality is questionable for all IPIP-VIA strengths (Table 1.6). On the basis of findings of parallel analysis, even essential unidimensionality seems out of reach for most of the item sets. Although not all additional factors indicated by parallel analysis may reflect substantive dimensions, the deviations from unidimensionality were severe enough to prevent several items from loading substantially on a common strength factor. Moreover, several of the 252 items from the IPIP-VIA item pool had low communalities in multi-factor models, including several items in the currently recommended 213-Item IPIP-VIA scales. Our findings question the item quality, scale compositions, and the appropriateness of using IPIP-VIA scale sums or scale means. While they supported our approach towards *shortening* the IPIP-VIA scales, these findings also limited our options for *balancing* the scales. To allow for balanced, content-valid and (essentially) unidimensional character strength measures, the IPIP-VIA scales needed to be refined.

Research Question 2:

Does the IPIP Comprise a Sufficient Number of Content-Valid Items per Strength Such That a Selection of Cross-Culturally Applicable Items Would Form Balanced-Keyed and (Essentially) Unidimensional Scales?

The deviations from unidimensionality identified in RQ1 call for a careful item selection to weed out problematic items. Yet, purely quantitative approaches to item selection (e.g., optimizing internal consistency) do not automatically ensure content validity: “If the content of an instrument cannot be defended with respect to the use of the instrument, construct validity cannot be obtained” (Sireci, 1998, p. 112). Five conditions must be satisfied to ensure content validity of a measure (Guion, 1977): The content domain is relevant to the purposes of measurement; the domain is unambiguously defined; the behavioral content has an accepted meaning or definition; qualified judges consensually agree that the domain has been adequately sampled; and the content must be observable and evaluated. Moreover, content validity in one culture or language does not automatically ensure cross-cultural applicability (i.e., culture fairness) of measures—which, however, is indispensable because the VIA framework conceives of character strengths as cultural universals (e.g., Church & Lonner, 1998).

Given the quality of items, and particularly the small number of negatively keyed items in IPIP-VIA, we were severely limited regarding the number of items available for creating balanced short scales. In selecting items from IPIP-VIA through expert ratings, we therefore pursued several interrelated goals in parallel: (1) Considerably shortening the scales to four items (i.e., 96 in total); (2) approximating unidimensionality; and optimizing (3) content validity and (4) the quality of item wording. We also paid heed to (5) cross-cultural applicability (or “fairness”). Finally, (6) the resulting item set should be balanced-keyed to control for acquiescence. In this way, we sought to address the previously discussed issues of IPIP-VIA items.

We used the expert ratings of all 252 IPIP-VIA items to select the best four items per strength. To become able to provide a sufficient number of content-valid—in particular reversed—items, our selection also considered IPIP items identified as potentially suitable for character strength assessment from other (non-VIA) IPIP sections. We computed indices of inter-rater agreement across all raters' judgments of item suitability. We then tested the unidimensionality of the selected item sets in four online samples (two from the UK and two from Germany) that were drawn specifically for this purpose. Because we had encountered some disagreement during the selection and because a few of the initial items and scales (subsequently termed Version A) were not fully satisfactory yet, we compiled Version B items and scales and collected similar evidence about unidimensionality (see Overview).

Method

Expert Ratings and Item Selection Criteria. To select items according to the six above-specified criteria, four professionals with expertise in personality/character assessment, scale development, and cross-cultural survey methodology screened the 252 IPIP-VIA items. They identified the best-suited 96 IPIP items (i.e., four items for each of 24 character strengths) that would serve as the basis for translation and adaptation from English to German. The experts selected the items based foremost on how well the item content/wording matched the definition of the target strength (see Table 1.1). They were instructed to prefer concise items in simple language (clarity), and to avoid items whose content fit more than one character strength (distinctness). Moreover, culture-specific connotations and idiomatic expressions were avoided to support cross-cultural equivalence. Each rater determined their preferred combination of the four most suitable items per scale (considered "essential"), with the option of marking one item per scale as "useful but not essential" (as a basis for establishing consensus). Ultimately balancing positive and negative items (acquiescence correction) required that discrepancies over which items to select had to be resolved by means of establishing post-negotiation consensus (see SOM_IPIP-VIA-R_Method_Item-Selection.pdf).

Based on the experts' inclusion/exclusion ratings, we computed the content validity index (CVI; Polit et al., 2007). For CVI, experts usually evaluate items according to four options, which are then collapsed into two categories: relevant or not relevant. In our case, the CVI at the item level (I-CVI) is directly computed as the proportion of "relevant"-judgments across raters; it reflects only the agreement about relevance (any agreement about non-relevance is not implied). Averaging I-CVI values of items selected into a scale yields CVI at the scale level (S-CVI). All CVI values range from 0 to 1. Common rules of thumb suggest eliminating items with I-CVI < .70, revising items with .70 < I-CVI < .80, and accepting items with I-CVI > .79. Others have suggested accepting items with I-CVI > .70, .78, or .80 (see Davis, 1992; Polit et al., 2007; Tilden et al., 1990). Yet others consider an average I-CVI as good (or excellent) content validity at the scale level if S-CVI > .80 (or .90; Polit et al., 2007). These rules of thumb are somewhat arbitrary and not universally applicable.[6] Given our aim to select an *a priori* fixed number of four items considered to represent the best selection of items from an existing (i.e., fixed) item set with a set of four raters, we favored a cutoff of S-CVI ≥ .75. To safeguard against unreliability of experts or chance agreement, we estimated inter-rater reliability (Hallgren, 2012).

Statistical Analysis of Closeness to Unidimensionality. We then tested the unidimensionality assumption for each strength scale based on the initially selected items (IPIP-VIA-R, Variant A) in two samples (Sample 2 from the UK and Sample 3 from Germany). Moreover, we compared the initially selected items to a slightly modified item set (IPIP-VIA-R, Variant B) in two independent samples from the same countries (Samples 4 and 5). Item descriptives collected with Samples 2–5 can be found in SOM_IPIP-VIA-R_RQ2_Descriptives.pdf.

Using the FACTOR software (Lorenzo-Seva & Ferrando, 2013), we ran an implementation of Horn's (1965) parallel analysis based on minimum rank factor analysis (MRFA; Timmerman & Lorenzo-Seva, 2011), which exploits the (complete-case) polychoric correlation matrix for short scales based on a mere four (ordinal) items per strength.[7] Again, 500 random correlation matrices, obtained through raw score permutation (Buja & Eyuboglu, 1992), suggested the number of necessary dimensions. We ran MRFA under the constraint of extracting two dimensions at the most (which would reflect a potential method factor for acquiescence besides the substantive strength factor).

FACTOR also provided Explained Common Variance (ECV; Rodriguez et al., 2016) as an index of unidimensionality. ECV expresses the percentage of variance attributable to the (first) factor common to all items (depending on the number of factors also termed the general factor). We inspected the so-called Mean of Item Residual Absolute Loadings (MIREAL), too, which expresses the average of the item loadings on a secondary factor (extracted with Promin rotation) and represents a general measure of departure from unidimensionality (Ferrando & Lorenzo-Seva, 2018). ECV indicates the dominance of the first MRFA factor. MIREAL depicts a potential residual (second) MRFA factor. ECV should be high (> .60), and MIREAL should be low (< .40), especially to prevent bias when using simple means or sum scores (see Table 1.5). The chosen cut-offs do reflect the small number of items, the non-redundant representations of strengths as well as the expected deviation from strict unidimensionality due to the balanced keying direction of items.

Results

Inter-rater Reliability and Content Validity. An initial inspection showed 100% spontaneous agreement among experts about excluding 82 (33%) of the 252 IPIP-VIA items. Another 53 (21%) items were judged "essential" by merely one rater. By contrast, experts consensually judged 28 (11%) items as suitable, and 43 (17%) were preferred by a three-quarter majority among the raters. For the remaining 46 (18%) items, there was a tie among the four raters. Never did we include an item into a strength scale if not at least half the experts considered it “essential” (or at least “useful” in one case—in the absence of a better reversed item for measuring Capacity for Love). With half the items unselected and only few reversed-keyed items in the pool, aiming at balanced four-item rather than six-item scales was the only realistic option with IPIP. Four items yield a reasonable compromise between reliability and efficiency for measuring strengths that reside at a similar level as personality facets, which the BFI-2 too assesses with four balanced items (Soto & John, 2017a).

We used the intraclass correlation coefficient to quantify the agreement of the experts on their preferred selections (Hallgren, 2012; Koo & Li, 2016). Using a two-way mixed effects model with raters as a fixed factor, the single-rater reliability was $ICC_{3,1} = .45$. The single-rater *ICC* matches Fleiss’s (1971) kappa ($\kappa = .45$) and indicates moderate above-chance agreement between pairs of raters according to standards provided by Landis and Koch (1977). The reliability of the average judgment across all raters was $ICC_{3,4} = .76$. Given that raters could opt for different yet equally viable four-item sets, we consider this degree of reliability highly satisfactory. For illustrative purposes, after forming *post-negotiation consensus* and *including non-VIA IPIP items* to form *balanced* four-item scales (yielding Version A of IPIP-VIA-R), we coded the respective items as showing “perfect rater agreement”, yielding agreement at $ICC_{3,1} = .54$ (single-rater), and $ICC_{3,4} = .83$ (average of raters).[8]

Thereafter, content validity ratings at the level of balanced scales (Variant A) amounted to an average S-CVI = .84 across 24 scales (range = .63–1.00). For seven scales, S-CVI exceeded the set threshold of .75, and for another 13 scales S-CVI exceeded even all scales’ average S-CVI (≥ .85). Although four scales (Curiosity, Kindness, Originality, and Perspective) contained items that we consider suboptimal and that lowered S-CVI values below the set bar, we accepted these negotiated scales tentatively as the best possible set of items that can be obtained within the IPIP. Yet, out of 24 strength scales, 20 scales had good or very good content validity. Thus, Variant A represented substantially improved scales over the IPIP-VIA scales currently in use, whose items cannot all be recommended with good conscience. Still, over the course of the negotiation process, while considering alternative translations, and in light of the obtained statistical indices for unidimensionality (see next) and reliability estimates (see RQ3), we reasoned that five viable alternative items in Variant B scales existed (and an additional five substitute translations) that we compared in the next step.

Unidimensionality. To evaluate the dimensionality of all 24 character strengths, we focused on Samples 4 and 5, which allow for a direct within-sample comparison of Variants A and B. (For readers interested in the detailed first findings for Variant A scales based on our Samples 2 and 3, we provide the respective columns in Table 1.6, too.) Parallel analysis clearly indicated lower dimensionality for all 24 short scales than for the longer item sets (see RQ1). In the UK (Sample 4), parallel analysis suggested a single dimension to underlie eight Variant A scales, and yet another three Variant B scales. For Citizenship/Teamwork and Kindness, two dimensions seemed possible in Sample 4 (but not so in Sample 2 before). In the remaining scales, not even a single factor could be reliably suggested after having opted for 95% statistical confidence in determining the likely number of factors with parallel analysis. In Germany (Sample 5), the majority of 19 Variant A and B scales supported the one-factor notion, while none of the German scales appeared to be two-dimensional. However, there was a peculiar finding for some scales (in Germany, for five Variant A scales and three Variant B scales; in the UK, for 14 Variant A scales and two Variant B scales), namely that the number of suggested dimensions fell below one, while these values fluctuated across countries and samples.

Given the fluctuating number of suggested factors, we focused on ECV and MIREAL in Samples 4 and 5, whose values seemed rather comparable across countries and samples (see Table 1.6). An ECV value close to one would suggest a very strong common factor and support an (at least essentially) unidimensional scale. Yet, with the built-in violation of unidimensionality (balanced item keying), an important question was if the secondary factor (method factor driven by acquiescence) would remain rather negligible. The common factor per scale explained reasonable percentages of variance, average ECV_{UK} = .67 and .68, average ECV_{DE} = .75 and .76, for scale Variants A and B, respectively. The mean item-residual average loading values showed the presence of a constrained source of common variance in each strength scale, average $MIREAL_{UK}$ = .39 and .39, average $MIREAL_{DE}$ = .30 and .29, for scale Variants A and B, respectively. Often lower MIREAL values tended to be supported by higher ECV values, usually favoring Variant B over Variant A scales, as intended. Yet, direct strength scale comparisons were sometimes ambiguous (e.g., Spirituality/Religiousness), not always consistent across countries (e.g., Appreciation of Beauty and Excellence), thus not always clearly supporting the superiority of one variant over the other. Therefore, we deferred the ultimate choice between the Variants A and B until we had inspected the fit of proper measurement models, factor loadings, and scale reliability (RQ3–4).

Discussion

Based on expert ratings that achieved high inter-rater reliability, we were able to shorten the IPIP-VIA scales by about 50% while concomitantly improving several measurement properties: unidimensionality, content validity, cross-cultural applicability, and balanced keying. Although measuring each character strength with four items was our aim from the outset (similar to the four items per facet in the BFI-2; Soto & John, 2017a), the expert ratings also suggested that the IPIP hardly comprised more than four items suitable for achieving balanced character strength scales. Instead, experts agreed that most of IPIP-VIA items analyzed in RQ1 did not represent their respective strength adequately or exhibited other issues (e.g., lack of cross-cultural applicability). Selecting suitable, balanced item sets for each strength scale, however, yielded respectable content validity (S-CVI ≥ .75 for 20 of 24 strengths).

Given the present scale lengths and sample sizes, MRFA-based parallel analysis was not trustworthy for determining the minimally required number of common factors (see Lim & Jahng, 2019, for an explanation of underestimating the number of factors in short scales). Our reasoning was supported by the comparably consistent findings for ECV and MIREAL values and the changes therein when amending items to form Variant B scales. Any scale achieved (at least essential) unidimensionality when judged by the reasonable percentages of common variance of underlying strength factors. Where the ECV-threshold of .60 was not passed by Variant A (Prudence and Appreciation of Beauty and Excellence in the UK), Variant B passed it. In general, MIREAL values showed that the scales guarded against severe violations of unidimensionality. Noteworthy, ECV values were higher and MIREAL values lower in the German compared to the UK samples—a finding that is, overall, in line with the pattern of parallel analysis. On average, the German item adaptations yielded closer approximation to unidimensional scales than the English source items. Therefore, we were satisfied with the obtained closeness to unidimensionality, especially since acquiescent responding as a secondary source of variance was not yet (statistically) controlled in these analyses (see RQ3).

Our finding that most existing VIA strength items fall short of accepted quality standards resembles the conclusions reached by Ng and colleagues (2017). These authors found that more than 50% of VIA-IS items had to be eliminated to achieve proper item loading patters for the 24 strengths. Their item selection was purely empirical and consisted in eliminating items with the poorest factor loadings iteratively. Our selection, by contrast, jointly optimized several substantive criteria. Together with further findings on the questionable representation of character strength cores by several items in extant VIA-IS scales (McGrath, 2014, 2019; McGrath & Wallace, 2021), our results from RQ1 and RQ2 caution against using unit-weighted scores to represent the VIA character strengths as measured with IPIP-VIA (and possibly with related measures of character strengths, too). Unit-weighted scores derived from these character strength measures are not unidimensional and contain several less-than-optimally worded items that threaten content validity and cross-cultural applicability. Together with previous VIA-IS-based findings, our IPIP-VIA-based findings must have repercussions for the trustworthiness of previous analyses of associations among character strengths, identification of hierarchical structures, reliability estimates, construct validity, and predictive validity, let alone measurement equivalence across cultures.

Research Question 3:

Do IPIP-VIA-R Measurement Models Achieve Good Fit?

We sought to establish measurement models for IPIP-VIA-R that concomitantly would support choosing between the scale Variants A and B. For most of the time in research on the VIA framework, developing appropriate, character-strength specific measurement models was a neglected aspect. Only recently, Ng et al. (2017), Feraco et al. (2022), McGrath et al. (2022), and McGrath (2022) looked at unidimensionality (issues) of full and abbreviated versions of VIA-IS and VIA-IS-R, respectively. Selecting suitable IPIP-VIA items in such a way that all 24 short scales were balanced-keyed (i.e., two positively and two negatively worded items) allowed us to develop reflective measurement models for each character strength in which we modeled each character strength as a unidimensional latent variable and additionally controlled for respondents' acquiescent response style (henceforth ARS; Aichholzer, 2014; Billiet & McClendon, 2000). As discussed earlier, ARS is a major source of bias in survey research, especially in research based on general population samples and in cross-cultural research (Lechner et al., 2019; Podsakoff et al., 2012; Rammstedt & Farmer, 2013). Unless accounted for by means of balanced item keying and appropriate specification of ARS in measurement models, individual differences in ARS can tarnish means and all covariance-based statistics of items and scales (McCrae et al., 2001). Consequently, failing to control for ARS also biases factor loadings, model fit statistics, and measurement invariance (e.g., De Beuckelaer et al., 2010; Schimmack et al., 2002).

One challenge in controlling for ARS is that unless the item set consists of perfectly balanced antonyms, an ARS factor may confound acquiescence with item-specific content. This is particularly true when only a few (e.g., 2 or 4) items are available from which to extract respondents' ARS. Fortunately, the validity of an ARS factor increases as the number of balanced and substantively heterogeneous items increases (Aichholzer, 2014; De Beuckelaer et al., 2010; Lechner et al., 2019). This is because ARS (i.e., agreement irrespective of item content and keying) is the only plausible source of indiscriminate agreement across a large and diverse set of items. We therefore capitalized on 96 balanced-keyed items in all IPIP-VIA-R scales to garner information about a respondent's ARS. Specifically, we computed a manifest ARS index as the ipsative (i.e., within-respondent) mean across all 96 balanced items. The ipsative mean scores across a large set of heterogeneous and balanced items reflects the systematic (dis-)agreement with items, irrespective of content. We then included this index as a control variable in the measurement model to remove ARS bias from the character strength indicators.

Figure 1.2 shows the tested CFA model for a given character strength. Four indicators reflected the latent character strength (identified with unit variance) with freely estimated factor loadings. Whereas strictly unidimensional models would deny the relevance of a method factor for balanced item sets, we controlled for acquiescence by regressing each item on the ARS index. Manifest ARS indices typically correlate very highly with latent ARS variables (Billiet & McClendon, 2000), yet specifying a secondary latent ARS variable for each strength would require longer strength scales. Because the variance of the ARS index is given and fixed, we estimated the regression weights *freely* while constraining them to equality across all four items of each scale. Hence, the impact of ARS can differ per strength scale (i.e., *scale-specific acquiescent responding*; Greenleaf, 1992; Schimmack et al., 2002; Weijters et al., 2010).

We used these novel measurement models to establish whether each of the 24 character strength short scales can be modeled as a unidimensional latent variable (after controlling for ARS) and to compare the scale Variants A and B in order to decide on the final versions of the IPIP-VIA-R strength scales. Moreover, the models provide the foundation for our subsequent analyses of the scales' reliability, validity, and measurement invariance (RQ4–7).

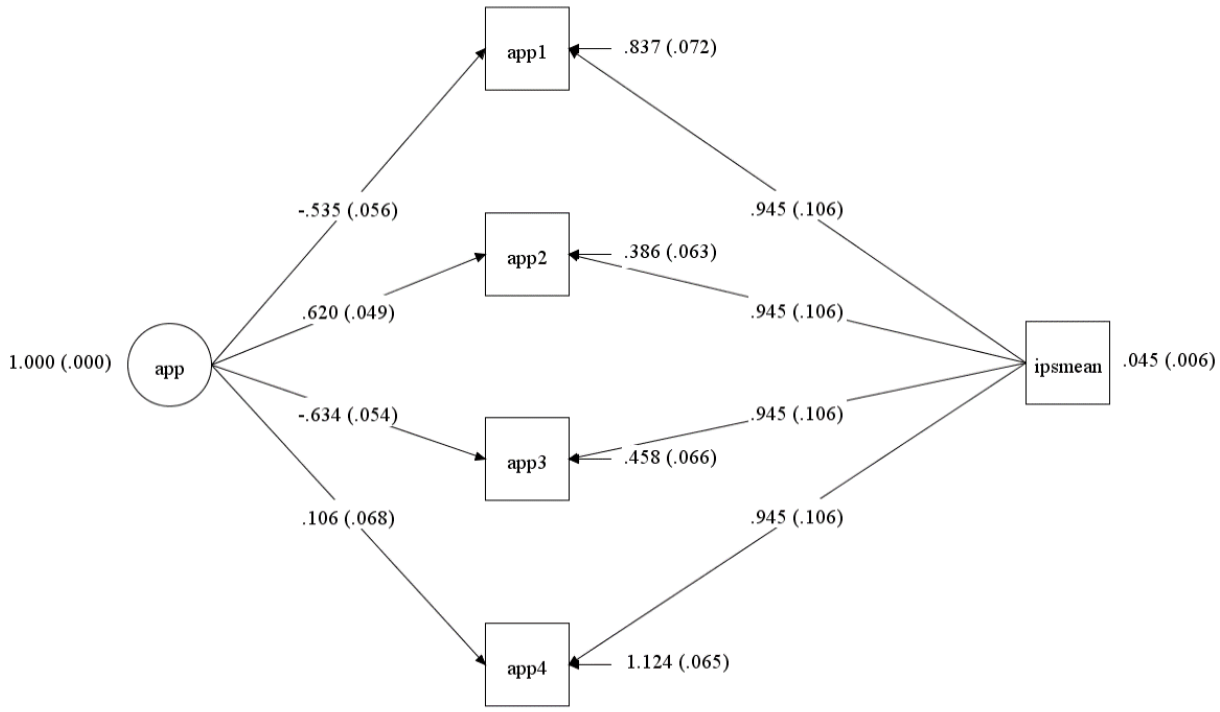


Figure 1.2. CFA measurement model for a common latent variable (here: Appreciation of Beauty and Excellence, variant A, depicted as 'app') based on four indicators (app1–app4). Acquiescence is controlled by a manifest covariate 'ipsmean' (i.e., the ipsative mean across 96 VIA items), which impacts on all indicators with the same weight irrespective of keying direction.

Method

We estimated CFA model parameters for Samples 2–5 with robust maximum likelihood estimation (MLR; Rhemtulla et al., 2012). We handled the few missing item responses with Full Information Maximum Likelihood (FIML) under the assumption of missingness at random (MAR). We judged the fit of the measurement models by the χ^2-test with *df* model degrees of freedom and the resulting goodness-of-fit indices. At large sample sizes, χ^2 will indicate statistically significant deviations of the model-implied from the empirical variance-covariance matrix even if misfit is trivial. In line with current guidelines (e.g., Kline, 2011), we therefore considered the largely sample-size independent Comparative Fit Index (CFI), the parsimony-rewarding Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). Previously suggested cutoff heuristics cannot be applied to very short scales and complex models (Marsh et al., 2004), especially with few *df* remaining and with low factor loadings (e.g., due to rather heterogeneous item content; Heene et al., 2011), as is partly true for the current strength short scales. Therefore, we considered models as tenable when CFI ≥ .90, RMSEA ≤ .10, and SRMR ≤ .08 (see Table 1.5).[9]

Results

Model Fit and Factor Loadings of IPIP-VIA-R Strength Scales: Variants A and B. We first summarize the results for Version A scales as obtained with Samples 2 and 3 (for detailed results, see Table B1 in SOM_IPIP-VIA-R_RQ3_Measurement-Model-Fit-&-Loadings.pdf), because these models, their respective fit indices, and their respective factor loadings constituted the statistical starting point for engaging with Version B scales. Across all 24 strengths, we obtained very good average values of CFI = .970 and .949, RMSEA = .056 and .072, and SRMR = .044 and .050 for the UK and Germany, respectively. According to the conservative (but non-significant) χ^2-test, eight scales even fit perfectly in both countries, and another five scales in the UK. Considering fit heuristics, all scales passed the SRMR cutoff, and nearly all scales passed the CFI threshold except for EQU, LOV, and SPI. According to RMSEA, which tends to be biased in short scales, 16 scales were suitable in both countries, yet eight scales flagged in at least one country. Because we obtained relatively poor CFI and RMSEA values across both countries, we developed a Variant B scale of LOV in which we exchanged item #4 (see SOM_IPIP-VIA-R_Method_Item-Selection.xlsx to retrace scale revisions to obtain Variant B scales).

In Samples 2 and 3, all factor loadings were in the expected direction and standardized loadings were, on average, appropriate for short scales covering character strengths in breadth. Absolute loadings across all 24 scales averaged to the following means of $\lambda(1–4)$ = .51–.60 in the UK and .54–.61 in Germany. The minimum and maximum λs encountered for any specific items were .10 (APP #4 in DE) and .83 (SPI #4 in DE). We reinspected eleven scales that provided at least one item with a low loading ($\lambda < .40$), which prompted us to develop Variant B scales for APP (exchange of item #4), CIT (revised translation of items #2 and #4), EQU (revised translation of item #4) , FOR (revised translation of item #1) , PRU (exchange of item #4) , SEL (revised translation of item #4) , SPI (exchange of item #2) , and ZES (exchange of item #2).

When comparing the model fit of Variant A scales in Samples 2 and 3 against those in Samples 4 and 5, the scale quality was largely similar. For instance, two Version A scales we had flagged (Love of Learning and Spirituality/Religiousness) had problematic model fit in all samples. Given the highly comparable findings, we deemed contrasting and assessing scale Variants A and B within Samples 4 and 5 useful (and potentially better than comparing Variant A from Samples 2 and 3 with Variant B from Samples 4 and 5). Regarding this within-sample comparison, the average fit across all Version A scales was good or very good, CFI = .975 and .964, RMSEA = .060 and .068, and SRMR = .042 and .055, for the UK and Germany, respectively. Similar good fit resulted, on average, when analyzing Version B scales, CFI = .980 and .966, RMSEA = .055 and .069, and SRMR = .041 and .058, for the UK and Germany, respectively (see Tables B2 and B3 in SOM_IPIP-VIA-R_RQ3_Measurement-Model-Fit-&-Loadings.pdf). With non-significant χ^2-tests, 12 and 9 Variant A scales as well as 13 and 8 Variant B scales obtained perfect fit in the UK and Germany, respectively. The remaining models had a χ^2-value conventionally labelled as statistically significant, but the model fit indices fell in a range conventionally labelled as *sufficient* to *very good*.

Considering jointly the resulting model fit and factor loadings, some Version B scales (e.g., Appreciation of Beauty and Excellence with nonsignificant χ^2_{UK}), but not all Version B scales (e.g., Forgiveness: nonsignificant χ^2_{DE} in Version A, but significant in Version B), worked better. Despite overall good model fit, the values did not consistently favor all Version B scales. Given the equivocal results regarding these scale variants, we based our decision about which variant to retain in the final version on both linguistic considerations and all empirical evidence on dimensionality (ECV, MIREAL) from RQ2 as well as the CFA model fit and factor loadings. We additionally considered scale reliability (presented next in RQ4). On this basis, we preferred three Variant B scales (APP, LOV, ZES) yet retained two Variant A scales (PRU, SPI) because the corresponding Variant B scales had several unacceptably low factor loadings in Germany. Of the four scales with either one (EQU, FOR, SEL) or two (CIT) alternative item translations, we always preferred Variant B. For details, see SOM_IPIP-VIA-R_RQ3_Item-scale-variant_Selection-Criteria.xlsx (and the accompanying documentation SOM_IPIP-VIA-R_RQ3_Item-scale-variant_Selection-Criteria.pdf). Table 1.7 presents the resulting final 96 items of the refined IPIP-VIA-R short scales.

Model Fit of Final IPIP-VIA-R Strength Scales. We provide the findings for the final IPIP-VIA-R scales in Table 1.8. For Germany, eight scales showed excellent CFA model fit without a statistically significant χ^2-deviation. The same was true for the same eight scales (plus another five scales) in the UK. Model fit fell in a range that would be conventionally labeled as *sufficient* to *very good*: Across all 24 strengths, we found average values of CFI = .980 and .965, RMSEA = .055 and .070, and SRMR = .041 and .059, for the UK and Germany, respectively. For all models with significant χ^2-test, except for Spirituality/Religiousness, the majority of the three fit indices fell within acceptable limits.

With the caveat that it might be inflated for short scales, RMSEA flagged the fit of the model for Hope as potentially problematic consistently in both countries. Similarly, RMSEA consistently leaned towards the upper boundary for the Leadership scale.

Discussion

These findings demonstrate that unidimensional measurement models with ARS control showed good fit for most of the 24 character strengths in four samples representing two countries and languages. This suggests that our attempts to obtain unidimensional measures of the character strengths by discarding items that fell short of conventional standards in terms of linguistic quality and content validity were successful. In line with the findings from RQ2, our results also attest to the importance of modeling ARS in order to obtain well-fitting models that provide a solid foundation for testing reliability, measurement equivalence, and construct validity.[10]

Only three of the measurement models (i.e., SPI, HOP, and LEA) did not achieve sufficiently good fit across all the fit indices and/or had at least one item with a lower factor loading. One might speculate that the limits of better measurement within the IPIP-VIA framework become tangible here. However, the models for these three strengths might still be improved by permitting correlated errors that may be expected when psychological tests are shortened, because with small numbers of items the error score estimates are based on increasingly insufficient true score estimates (Beauducel & Leue, 2014). We will explore this possibility before testing cross-cultural equivalence (RQ5) because, rather than simply optimizing fit within countries, any model adjustment should be applicable across countries.

Table 1.7

Original English and German Adaptations of IPIP-VIA-R Items [including (revised) Scale Labels]

Pos.	Key	English Items (IPIP Variable No.)	German Items
		Appreciation of Beauty and Excellence [APP]	**Sinn für das Schöne und Exzellenz [APP]**
4	−	Am rarely aware of the natural beauty in the environment. (v208)	Ich mache mir die Schönheit der Natur nur selten bewusst.
28	+	Experience deep emotions when I see beautiful things. (v102)	Wenn ich etwas Schönes sehe, bewegt mich das sehr.
52	−	Fail to notice beauty until other comment on it. (v139)	Ich bemerke Schönheit erst dann, wenn andere darauf aufmerksam machen.
76	+	Feel it's important to live in a world of beauty. (v52)	Ich spüre, dass es wichtig ist, in einer Welt voller Schönheit zu leben.
		(Capacity for) Love [CAP] - rev: Love [LOV]	**Liebesfähigkeit [CAP] - rev: Fähigkeit zu lieben und geliebt zu werden [LOV]**
3	−	Do not easily share my feelings with others. (v13)	Es fällt mir schwer, anderen meine Gefühle zu zeigen.
27	+	Know that there are people in my life who care as much for me as for themselves. (v55)	Ich bin mir gewiss, dass es Menschen in meinem Leben gibt, denen mein Wohl genauso wichtig ist wie ihr eigenes.
51	+	Can express love to someone else. (v234)	Ich bin fähig, meine Zuneigung anderen gegenüber auszudrücken.
75	−	Have difficulty accepting love from anyone. (v340)	Es fällt mir grundsätzlich schwer, Liebe anzunehmen.
		Citizenship [CIT] - rev: Teamwork [TEA]	**Soziale Verantwortlichkeit [CIT] – rev: Teamfähigkeit [TEA]**
6	−	Am not good at working with a group. (v16)	Ich kann einfach nicht gut mit anderen zusammenarbeiten.
30	+	Am an extremely loyal person. (v92)	Ich bin sehr loyal, verlässlich.
54	+	Support my teammates or fellow group members. (v236)	Ich setze mich für mein Team oder eine größere Gemeinschaft ein.
78	−	Prefer to do everything alone. (v73)	Ich mache lieber alles alleine.
		Curiosity [CUR]	**Neugier [CUR]**
7	−	Am not all that curious about the world. (v115)	Das Meiste in dieser Welt weckt nicht gerade meine Neugier.
31	+	Am excited by many different activities. (v151)	Ich kann mich für viele verschiedene Aktivitäten begeistern.
55	+	Can find something of interest in any situation. (v268)	Ich kann in jeder Situation etwas finden, was mein Interesse weckt.
79	−	Have few interests. (v342)	Es gibt nur wenige Dinge, die mich wirklich interessieren.

(continues)

Pos.	Key	English Items (IPIP Variable No.)	German Items
		Equity [EQU]	**Fairness [EQU]**
12	+	Treat all people equally. (v94)	Ich behandle alle Menschen gleich fair.
36	−	Take advantage of others. (v109)	Ich tue immer wieder mal Dinge auf Kosten anderer.
60	+	Believe that everyone's rights are equally important. (v120)	Mir ist es wichtig, dass alle Menschen die gleichen Rechte haben.
84	−	Treat others differently if I don't like them. (v241)	Wenn ich Leute nicht mag, behandele ich sie schlechter.
		Forgiveness [FOR]	**Vergebungsbereitschaft [FOR]**
5	+	Try to respond with understanding when someone treats me badly. (v269)	Ich bin nicht nachtragend, wenn mich jemand schlecht behandelt.
29	−	Hold grudges. (v72)	Es dauert lange, bis ich verzeihen kann.
53	+	Allow others to make a fresh start. (v149)	Ich gebe anderen immer die Möglichkeit für einen Neuanfang.
77	−	Find it hard to forgive others. (X210)	Ich kann anderen nicht so leicht vergeben.
		Gratitude [GRA]	**Dankbarkeit [GRA]**
10	+	Express my thanks to those who care about me. (v26)	Menschen, die sich um mein Wohl sorgen, zeige ich stets meine Dankbarkeit.
34	+	Am an extremely grateful person. (v175)	Ich bin von Dankbarkeit erfüllt.
58	−	Feel no gratitude to others. (Q58)	Anderen gegenüber kann ich echt keine Dankbarkeit empfinden.
82	−	Find few things in my life to be grateful for. (v248)	Ich finde in meinem Leben nur wenige Dinge, für die ich dankbar sein könnte.
		Hope [HOP]	**Hoffnung [HOP]**
21	−	Expect the worst. (v36)	Ich neige dazu, mit dem Schlimmsten zu rechnen.
45	+	Can find the positive in what seems negative to others. (v65)	Ich kann Gutes selbst dort finden, wo andere nur Schlechtes sehen.
69	+	Remain hopeful despite challenges. (v106)	Selbst bei Herausforderungen bleibe ich hoffnungsvoll.
93	−	Often think about the possibility of negative outcomes that are not likely to occur. (D47)	Ich male mir häufig schlimme Dinge aus, die wahrscheinlich nie passieren werden.

(continues)

Pos.	Key	English Items (IPIP Variable No.)	German Items
		Humor [HUM]	**Humor [HUM]**
19	−	Am not known for my sense of humor. (v45)	Ich bin nicht gerade für meinen Sinn für Humor bekannt.
43	+	Use laughter to brighten the days of others. (v111)	Ich heitere andere durch Lachen auf.
67	−	Am not fun to be with. (v76)	Ich bin keine besonders unterhaltsame Person.
91	+	Keep my sense of humor even in gloomy situations. (v216)	Selbst in trostlosen Situationen behalte ich meinen Sinn für Humor.
		Industry [IND] - rev: Perseverance [PEV]	**Ausdauer [IND] – rev: [PEV]**
8	+	Don't quit a task before it is finished. (v12)	Ich bleibe an Aufgaben solange dran, bis ich sie erledigt habe.
32	−	Don't finish what I start. (v49)	Ich fange viele Dinge an, bringe sie aber nicht zu Ende.
56	+	Finish things despite obstacles in the way. (v126)	Ich bringe Dinge zu Ende, auch wenn ich dafür Hindernisse überwinden muss.
80	−	Give up easily. (v200)	Oft gebe ich zu schnell auf.
		Integrity/Honesty [INT]	**Integrität/Ehrlichkeit [INT]**
15	+	Am trusted to keep secrets. (v10)	Geheimnisse, die man mir anvertraut, behalte ich für mich.
39	+	Keep my promises. (v14)	Ich halte meine Versprechen.
63	−	Lie to get myself out of trouble. (v143)	Wenn ich mir dadurch Ärger ersparen kann, nehme ich es mit der Wahrheit nicht so genau.
87	−	Cheat on people who have trusted me. (Q22)	Ich missbrauche das Vertrauen anderer.
		Judgment [JUD]	**Urteilsvermögen [JUD]**
13	−	Don't tend to think things through critically. (v118)	Ich neige nicht dazu, Dinge zu durchdenken und kritisch zu hinterfragen.
37	−	Don't think about different possibilities when making decisions. (v25)	Ich denke nicht groß über mögliche Alternativen nach, wenn ich Entscheidungen treffe.
61	+	Weigh the pros and the cons. (v257)	Ich wäge immer ab, was für eine Sache spricht und was dagegen.
85	+	Am valued by my friends for my good judgment. (v333)	Meine Freunde schätzen mich für mein gutes Urteilsvermögen.

(continues)

Pos.	Key	English Items (IPIP Variable No.)	German Items
		Kindness [KIN]	**Freundlichkeit [KIN]**
9	−	Get impatient when others talk to me about their problems. (v6)	Ich werde schnell ungeduldig, wenn andere mir von ihren Problemen erzählen.
33	+	Am never too busy to help a friend. (v24)	Um einem Freund zu helfen, finde ich immer irgendwie Zeit.
57	+	Go out of my way to cheer up people who appear down. (v88)	Ich scheue keine Mühen, um Menschen aufzumuntern, die niedergeschlagen wirken.
81	−	Am only kind to others if they have been kind to me. (v195)	Ich bin nur dann gut zu anderen, wenn sie auch gut zu mir sind.
		Leadership [LEA]	**Führungsvermögen [LEA]**
16	−	Have difficulty getting others to work together. (v54)	Ich bin nicht gut darin, andere zur Zusammenarbeit zu motivieren.
40	+	Am good at helping people work well together. (v96)	Ich bin gut darin, andere zur erfolgreichen Zusammenarbeit anzuleiten.
64	−	Am not good at taking charge of a group. (v105)	Ich bin nicht gut darin, in einer Gruppe die Führung zu übernehmen.
88	+	Am told that I am a strong but fair leader. (v329)	Andere nehmen mich zwar als sehr bestimmende, aber gerechte Führungsperson wahr.
		(Love of) Learning [LOV] - rev: Learning [LER]	**[Liebe zum] Lernen [LOV] - rev: Wissensdrang [LER]**
17	−	Don't like to learn new things. (v38)	Es macht mir keinen Spaß, Neues zu lernen.
41	−	Don't read nonfiction books for fun. (v332)	Sachbücher lese ich nur, wenn ich muss.
65	+	Am a true life-long learner. (140)	Ich bin jemand, der wirklich sein ganzes Leben hinzulernen will.
89	+	Am thrilled when I learn something new. (v75)	Ich bin begeistert, wann immer es etwas Neues zu lernen gibt.
		Modesty [MOD]	**Bescheidenheit [MOD]**
11	−	Like to stand out in a crowd. (v68)	Ich mag es, aus der Masse hervorzustechen.
35	+	Don't brag about my accomplishments. (v145)	Ich prahle nie mit dem, was ich erreicht habe.
59	−	Like to talk about myself. (v214)	Ich spreche gerne über mich selbst.
83	+	Would never be described as arrogant. (v321)	Niemand würde mich als arrogant bezeichnen.

(continues)

Pos.	Key	English Items (IPIP Variable No.)	German Items
		Originality/Creativity [ORI]	**Originalität/Kreativität [ORI]**
1	−	Am not considered to have new and different ideas. (v189)	Ich gelte nicht gerade als einfallsreich und originell.
25	+	Come up with new ways to do things. (v155)	Ich komme auf gute Ideen, wie man etwas auch ganz anders machen kann.
49	+	Am an original thinker. (v226)	Ich bin ein kreativer Kopf.
73	−	Have no special urge to do something original. (v335)	Ich habe keinen besonderen Drang, etwas Originelles zu machen.
		Perspective [PER]	**Weitsicht [PER]**
22	−	Am not good at figuring out what really matters. (v192)	Ich bin nicht gut darin zu erkennen, worauf es wirklich ankommt im Leben.
46	+	Have a mature view on life. (v225)	Ich habe eine sehr reife Sichtweise auf das Leben.
70	−	Am rarely consulted for advice by others. (v262)	Andere suchen nur selten meinen Rat.
94	+	Am considered to be a wise person. (v339)	Ich gelte als klug und weise.
		Prudence [PRU]	**Umsicht [PRU]**
23	+	Believe it is always better to be safe than sorry. (v62)	Ich lebe nach dem Motto "Vorsicht ist besser als Nachsicht".
47	−	Act before thinking through the consequences. (v207)	Ich handle ohne Rücksicht auf die Konsequenzen.
71	−	Like taking risks. (v277)	Ich gehe gerne Risiken ein.
95	+	Make careful choices. (v311)	Bei Entscheidungen gehe ich gerne auf Nummer sicher.
		Self-Control [SEL] - rev: Self-Regulation [REG]	**Selbstkontrolle [SEL] - rev: Selbstregulation [REG]**
20	+	Am a highly disciplined person. (v98)	Ich habe eine sehr gute Selbstbeherrschung.
44	+	Forego things that are bad for me in the long run even if they make me feel good in the short run. (v170)	Ich verzichte auf Dinge, die mir langfristig schaden, auch wenn sie sich kurzfristig gut anfühlen.
68	−	Let myself be taken over by urges to spend or eat too much. (v224)	Manchmal erliege ich der Versuchung, zu viel Geld auszugeben oder zu viel zu essen.
92	−	Give in to my urges. (v329)	Meinem Verlangen gebe ich - so gut wie immer - nach.

(continues)

Pos.	Key	English Items (IPIP Variable No.)	German Items
		Social/Personal Intelligence [SOC] - rev: Social Intelligence [SIQ]	**Soziale Intelligenz [SOC] - rev: Soziale Intelligenz [SIQ]**
14	−	Don't know how to handle myself in a new social situation. (v43)	Ich finde mich nicht gut in zwischenmenschlichen Situationen zurecht, die mir neu und unvertraut sind.
38	−	Have trouble guessing how others will react. (P405)	Ich bin nicht gut darin, die Reaktionen anderer Leute vorherzusehen.
62	+	Am good at sensing what others are feeling. (v225)	Ich habe ein feines Gespür dafür, was in anderen Menschen vorgeht.
86	+	Know what to say to make people feel good. (v295)	Ich weiß, was ich sagen muss, damit andere Menschen sich gut fühlen.
		Spirituality/Religiousness [SPI]	**Spiritualität/Religiosität [SPI]**
24	+	Am a spiritual person. (v46)	Ich bin ein spiritueller oder gläubiger Mensch.
48	−	Feel that life has no meaning. (Q215)	Ich glaube, dass das Leben keinen tieferen Sinn hat.
72	+	Believe that each person has a purpose in life. (v280)	Ich glaube, dass jeder Mensch eine Bestimmung im Leben hat.
96	−	Do not believe in a universal power or a God. (v282)	Ich glaube nicht an einen Gott oder eine höhere Macht.
		Valor/Bravery [VAL]	**Mut/Tapferkeit [VAL]**
2	+	Have taken frequent stands in the face of strong opposition. (v11)	Ich nehme häufig eine klare Haltung ein, auch gegen starken Widerstand.
26	−	Do not stand up for my beliefs. (v53)	Ich stehe nicht für meine Überzeugungen ein.
50	+	Don't hesitate to express an unpopular opinion. (v82)	Ich schrecke nicht davor zurück, eine unbeliebte Meinung zu vertreten.
74	−	Don't speak my mind freely when there might be negative results. (v130)	Wenn es negative Folgen haben könnte, meine Meinung zu äußern, dann halte ich sie lieber zurück.
		Zest/Vitality [ZES]	**Elan/Tatendrang [ZES]**
18	+	Awaken with a sense of excitement about the day's possibilities. (v326)	Ich erwache mit Vorfreude auf das, was mir der Tag bringt.
42	−	Am described as grumpy. (v32)	Ich werde oft als mürrisch, schlecht gelaunt beschrieben.
66	+	Look forward to each new day. (v182)	Ich freue mich auf jeden neuen Tag.
90	−	Don't have much energy.	Ich habe nur wenig Energie.

Note. Pos. = Position of item in IPIP-VIA-R questionnaire; Key = keying direction or agreement/disagreement with an item, scored 1 to 5 (+) or 5 to 1 (−); rev. = revised scale labels (to reflect the core definition of each strength and the content of each balanced item set better) together with revised abbreviations for IPIP-VIA-R scales, which may differ from the seminal IPIP-VIA scale names and abbreviations to avoid confusion (e.g., LOV has previously referred to Love of *Learning*, and not Capacity for *Love*, which has been addressed as CAP etc.).

Table 1.8

CFA Model Fit & Factor Loadings: IPIP-VIA-R – Final IPIP-VIA-R Scales (Samples 4 and 5)

	χ^2_{SB} (df = 5)		P-Value		CFI		RMSEA		SRMR		$\lvert\lambda_1\rvert$		$\lvert\lambda_2\rvert$		$\lvert\lambda_3\rvert$		$\lvert\lambda_4\rvert$	
	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE
APP_B	8.54	6.00	.13	.31	.99	1.00	.04	.02	.04	.04	.37	.53	.54	.53	.64	.63	.56	.45
CAP	12.55	24.52	.03	.00	.98	.94	.06	.09	.04	.06	.42	.57	.46	.36	.60	.67	.77	.64
CIT_B	9.49	25.22	.09	.00	.99	.90	.05	.11	.04	.11	.53	.58	.45	.46	.78	.53	.41	.45
CUR	5.24	10.30	.39	.07	1.00	.99	.01	.06	.04	.05	.42	.53	.66	.72	.51	.54	.56	.78
EQU_B	6.60	8.30	.25	.14	1.00	.99	.03	.05	.03	.05	.66	.73	.41	.43	.77	.60	.47	.62
FOR_B	12.47	15.73	.03	.01	.99	.98	.06	.07	.05	.07	.37	.55	.68	.75	.41	.41	.67	.84
GRA	7.83	26.71	.17	.00	1.00	.93	.04	.10	.04	.10	.70	.45	.65	.61	.61	.55	.41	.62
HOP	53.67	41.92	.00	.00	.95	.93	.12	.12	.05	.06	.72	.71	.52	.50	.57	.55	.48	.63
HUM	10.00	5.60	.08	.35	.99	1.00	.05	.02	.03	.03	.41	.64	.69	.72	.61	.72	.59	.57
IND	3.71	16.79	.59	.00	1.00	.97	.00	.08	.02	.06	.41	.53	.63	.75	.73	.72	.63	.60
INT	9.67	13.84	.09	.02	.99	.97	.05	.08	.05	.10	.65	.66	.68	.69	.51	.47	.59	.57
JUD	11.87	11.97	.04	.04	.99	.98	.06	.05	.04	.06	.48	.39	.60	.55	.62	.61	.42	.45
KIN	11.56	16.55	.04	.01	.99	.95	.05	.08	.04	.08	.62	.58	.57	.53	.61	.59	.43	.36
LEA	25.60	27.76	.00	.00	.97	.95	.09	.09	.05	.04	.53	.57	.67	.72	.59	.56	.41	.55
LOV_B	12.49	6.62	.03	.25	.99	1.00	.06	.03	.05	.03	.70	.45	.30	.33	.56	.79	.67	.76
MOD	28.38	14.58	.00	.01	.95	.95	.10	.07	.05	.05	.42	.37	.56	.67	.57	.58	.49	.38
ORI	27.88	6.84	.00	.23	.97	1.00	.09	.03	.05	.03	.44	.60	.62	.61	.65	.74	.58	.59
PER	5.94	7.17	.31	.21	1.00	.99	.02	.03	.04	.04	.64	.49	.50	.66	.48	.49	.55	.59
PRU_A	9.73	22.16	.08	.00	.99	.96	.04	.09	.04	.07	.60	.72	.31	.33	.48	.65	.46	.59
SEL_B	8.96	13.12	.11	.02	.99	.97	.04	.05	.04	.05	.31	.35	.25	.45	.70	.55	.61	.42
SOC	15.93	5.94	.01	.31	.98	1.00	.07	.02	.05	.03	.51	.31	.61	.55	.64	.72	.61	.53
SPI_A	82.58	52.97	.00	.00	.84	.89	.17	.16	.08	.07	.39	.79	.61	.44	.71	.38	.37	.80
VAL	2.60	20.40	.76	.00	1.00	.96	.00	.09	.03	.07	.43	.64	.43	.42	.61	.75	.61	.52
ZES_B	8.51	17.70	.13	.00	1.00	.97	.04	.08	.03	.07	.67	.76	.37	.45	.76	.75	.55	.57
Mean	-	-	.14	.08	.98	.96	.06	.07	.04	.06	.52	.56	.53	.55	.61	.61	.54	.58

Note. N = 950 (474 and 476 in UK and DE, respectively). The presence of a scale name index indicates the item/scale variant tested. CFI = (robust) Comparative Fit Index; RMSEA = (robust) Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; $\lvert\lambda_1\rvert$ … $\lvert\lambda_4\rvert$ = factor loadings (standardized regression weights), with loading sign reversed for negatively keyed items.

Research Question 4:

How Reliable are the Final IPIP-VIA-R Strength Scales?

Although the latent measurement models established in RQ3 showed adequate fit and constitute an optimal way of modeling character strengths as part of structural equation models (SEM), in practice many researchers will use manifest scale scores for character strengths. When manifest scale scores (as opposed to latent-variable modeling or plausible values) are used, measurement precision is an essential consideration. Next, we therefore shed light on the reliability of the 24 final IPIP-VIA-R short scales. Because "the choice of a single statistic to summarize the accuracy of an instrument is not the best report that can be made" (Cronbach & Shavelson, 2004, p. 414), we computed several different reliability indices (for index definitions and cutoff heuristics, see Table 1.5). Our focus was on internal reliability indices that are based on the appropriate measurement models specified in RQ3 (i.e., omega). For comparability with prior research, we also report more traditional reliability estimates (i.e., Cronbach's alpha), even though we note that some of these reliability estimates do not do justice to the measurement models for the character strengths. Finally, we also computed test-retest reliability. We computed each reliability estimate separately for each country. We present a summary evaluation of the final IPIP-VIA-R item sets (Samples 4 and 5; for scale-specific details, see Table 1.9; for statistical details and reliability outcomes in Samples 2 and 3, see SOM_IPIP-VIA-R_RQ4_Reliability.pdf).

What is considered "acceptable" or "sufficient" reliability depends on the research purpose. Population surveys often require cost-effective short scales, resulting in a trade-off between saving resources and aiming for high psychometric quality (e.g., Kemper et al., 2019). Whereas longer, more reliable scales are required for precise individual diagnostics, short surveys often compensate lower reliability estimates by means of larger sample sizes (Rammstedt et al., 2021). Also note that reliability is study-specific, as it depicts how reliable the individual differences are in a given sample of participants. Cutoff criteria for evaluating reliability need to reflect the different purposes of measurement (Nunnally & Bernstein, 1994). Short scales, especially designed to preserve the substantive breadth of the target constructs, often have reliabilities that fall below conventional thresholds; despite their lower reliability, they often predict life outcomes almost as well as longer scales (e.g., Rammstedt et al., 2021; Soto & John, 2019; Thalmayer et al., 2011). Considering the brevity of the IPIP-VIA-R scales, we therefore accepted reliabilities of $.60 < \rho < .70$ as sufficient for most research purposes, $.70 < \rho < .80$ as good, and values $\rho \geq .80$ as excellent.

Method

We computed the following reliability indices. Irrespective of the specific estimate, reliability (ρ) according to classical test theory reflects the proportion of true score variance in the observed score variance:

$$\rho = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\varepsilon^2}$$

Cronbach's Alpha (α). Alpha is the most widely used measure of internal consistency. For this reason, we report it here. However, coefficient alpha provides merely a lower-bound estimate for the reliability of the unit-weighted scale score. Furthermore, it is based on assumptions that are regularly violated in personality assessment, namely, strict unidimensionality, absence of residual correlations, and an essentially tau-equivalent model (i.e., identical factor loadings for all items; Cho & Kim, 2015; J. M. Graham, 2006). When these assumptions are violated, α potentially underestimates or, more rarely, overestimates scale reliability (Raykov, 1997, 1998, 2001; Sijtsma, 2009). We computed α for each scale after recoding the reversed-keyed items.

McDonald's Omega (ω). A reliability estimate that is more appropriate for tau-congeneric measures—and hence for most personality measures—is McDonald's Omega (McDonald, 1999). Different from α, ω does not assume tau-equivalence and can be generalized to handle violations of unidimensionality, providing a more *realistic* reliability estimate (e.g., Dunn et al., 2014). We estimated ω based on the accepted measurement models developed in RQ3 via Raykov's (1997) phantom-factor method, while accommodating the categorical nature of the ordinal indicators by a robust WLS estimator (WLSMV; B. O. Muthén et al., 1997; for details, see SOM_IPIP-VIA-R_RQ4_Reliability.pdf). This provides more realistic and trustworthy reliability estimates than coefficient alpha.

Test-Retest Reliability (r_{tt}). Test-retest reliability provides an estimate of the replicability of test scores when a test is administered twice. We estimated r_{tt} as the linear correlation between each strength, measured at two time points (Guttman, 1945). Different from α and ω, r_{tt} reflects the temporal consistency, that is, the reliability of all temporally stable sources of variance in the scale score (i.e., true score variance but also other stable influences such as acquiescent response style—which is, however, largely subtracted or averaged out in scores of balanced-keyed scales). For estimating the test-retest reliability r_{tt} of the 24 character strength scales, we re-invited participants from samples 4 and 5. The retest surveys consisted exclusively of the strength scales. The median time-lag amounted to 20 days (M_{DE} = 19.76, SD = 1.30; M_{UK} = 20.50, SD = 1.13). After quality filtering (and listwise deletion), the analytical retest sample rested on n = 224 (n_{DE} = 117, n_{UK} = 107; originally 244 participants with N_{DE} = 124 and N_{UK} = 120). The time-lag of three weeks represents a good trade-off between incurring inflated reliability due to memory effects and underestimation due to trait volatility.

We also computed the correlation between the latent strength variables across the two measurement occasions to provide a glimpse at the stability of reliable true score variance (which includes trait variance, but also stable state variance across three weeks). With its help it is easier to interpret the test-retest reliability coefficients, which entail true trait changes and state fluctuations across time, and which are also attenuated by measurement error. The retest correlation between the latent strength variables represents a natural boundary for the retest reliability of manifest scale scores.

Additional Reliability Coefficients. In addition to our three main reliability coefficients, we computed four additional coefficients: *Average Inter-Item Correlation (AIC), Average Variance Extracted (AVE), Factor Determinacy Index (FDI,* i.e., factor reliability*), and generalized H-Index* (gHI, i.e., construct replicability). These indices provide complementary information and may be of interest to some researchers and in specific research scenarios. Descriptions and results for these additional indices are also available in SOM_IPIP-VIA-R_RQ4_Reliability.pdf.

Results

Cronbach's Alpha (α). Average α-values were M_{UK} = .58 (SD_{UK} = .056) and M_{DE} = .63 (SD_{DE} = .075). Though some scales fared better than average (e.g., Humor, Industry/Perseverance), these values include "disappointing" coefficients even from the perspective of social surveys. However, apart from looking at short scales, α is likely to *under*estimate the reliability of IPIP-VIA-R because these scales do not meet α's restrictive assumptions as noted earlier.

McDonald's Omega (ω). As expected, ω was consistently higher than α and reached adequate levels, M_{UK} = .76 (SD_{UK} = .047) and M_{DE} = .75 (SD_{DE} = .066). 13 of the strength scales had even highly satisfactory values (ω ≥ .80) in either the UK or the German sample (four thereof in both countries), so that substantive strength variance represented more than three quarters of the total variance in the unit-weighted composite score.[11] These findings suggest that the primary factor in each IPIP-VIA-R scale is substantive, but this fact is concealed by α.

Test-Retest Reliability (r_{tt}). The average r_{tt} for manifest scale scores was M_{UK} = .66 (SD_{UK} = .076) and M_{DE} = .74 (SD_{DE} = .066). This is a realistic picture about what reliability to expect from manifest IPIP-VIA-R short scales (see McCrae et al., 2011). The German values are in line with the ω estimates, whereas r_{tt} in the UK was often lower than corresponding ω values. Note that even after accounting for measurement error, the correlations for the latent variables fell clearly below 1.00 (in nearly all cases), with M_{UK} = .83 (SD_{UK} = .105) remaining lower than M_{DE} = .90 (SD_{DE} = .057). The discrepant UK findings might be attributed to worse item wordings than we achieved for the German adaptations, alternatively to lower sample quality or other reasons responsible for higher state fluctuations than in Germany, unless one is willing to speculate on a theoretical level about less stable character strengths in the UK (i.e., culture-specific reasons) or unless short-term historical effects affect one country but not the other (i.e., political ruptures).

Table 1.9

IPIP-VIA-R Reliability: Final Scales (Samples 4 and 5)

	AIC		AVE		FDI		gHI		Alpha		Omega		$\text{Retest}_{\text{MANIFEST}}$		$\text{Retest}_{\text{LATENT}}$	
	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE
APP_{B}	.23	.27	.40	.34	.93	.85	.79	.67	.54	.58	.75	.69	.64	.74	.77	.90
CAP	.29	.29	.41	.38	.89	.88	.74	.73	.61	.62	.75	.73	.69	.77	.86	.94
CIT_{B}	.25	.23	.41	.32	.98	.85	.83	.66	.57	.54	.74	.68	.66	.63	.82	.94
CUR	.24	.39	.43	.50	.88	.94	.72	.82	.56	.72	.75	.82	.64	.80	.80	.93
EQU_{B}	.30	.33	.47	.42	.91	.91	.76	.77	.63	.67	.80	.78	.58	.71	.73	.83
FOR_{B}	.24	.37	.45	.52	.95	.94	.81	.84	.56	.71	.75	.83	.54	.75	.71	.84
GRA	.31	.28	.48	.38	1.00	.87	.83	.72	.63	.61	.84	.74	.65	.67	.73	.81
HOP	.28	.32	.48	.45	.92	.91	.79	.78	.60	.65	.80	.79	.72	.84	.91	1.00
HUM	.28	.42	.45	.50	.93	.91	.80	.79	.61	.74	.79	.83	.67	.82	.85	.92
IND	.32	.40	.43	.48	.91	.90	.78	.76	.63	.73	.81	.82	.53	.77	.61	.89
INT	.34	.34	.47	.39	.92	.89	.73	.69	.67	.65	.83	.80	.68	.77	.80	.95
JUD	.24	.22	.41	.31	.88	.83	.72	.66	.55	.51	.75	.67	.59	.58	.77	.82
KIN	.27	.23	.43	.33	.89	.84	.74	.67	.59	.54	.77	.69	.66	.75	.88	.95
LEA	.25	.34	.45	.40	.91	.90	.77	.77	.57	.67	.76	.76	.64	.75	.85	.94
LOV_{B}	.26	.31	.42	.39	.95	.94	.80	.83	.55	.62	.80	.81	.53	.71	.73	.82
MOD	.23	.24	.35	.30	.89	.86	.74	.69	.54	.55	.72	.66	.76	.76	.92	.88
ORI	.29	.38	.45	.47	.90	.90	.78	.77	.62	.71	.79	.80	.72	.79	.96	.95
PER	.25	.28	.44	.41	.87	.86	.71	.70	.56	.60	.74	.74	.73	.70	.94	.88
PRU_{A}	.17	.30	.35	.43	.90	.93	.74	.80	.45	.63	.63	.77	.76	.76	1.00	.90
SEL_{B}	.15	.17	.42	.27	1.00	.80	.89	.60	.42	.45	.68	.57	.53	.64	.61	.95
SOC	.31	.25	.47	.36	.93	1.00	.80	.87	.63	.56	.81	.72	.69	.66	.84	.80
SPI_{A}	.26	.37	.34	.48	.91	.95	.77	.83	.57	.70	.73	.81	.77	.83	.99	1.00
VAL	.22	.32	.42	.39	.88	.89	.74	.73	.54	.63	.73	.77	.75	.73	.90	.96
ZES_{B}	.29	.38	.48	.48	1.00	1.00	.89	.87	.62	.71	.81	.83	.72	.82	.88	.94
Mean	**.26**	**.31**	**.43**	**.40**	**.92**	**.90**	**.78**	**.75**	**.58**	**.63**	**.76**	**.75**	**.66**	**.74**	**.83**	**.90**

Note. N = 950 (474 and 476 in UK and DE, respectively), retest-n = 224 (120 + 124, respectively). The presence of a scale name index indicates the item/scale variant tested (in the UK, for CIT, EQU, FOR, SEL the A-versions were used). AIC = Average Inter-item Correlation; AVE = Average Variance Extracted; FDI = Factor Determinacy Index; gHI = generalized H-Index based on observed ordered categorical responses; Alpha = Cronbach's Alpha (internal consistency); Omega = Composite reliability for the common factor; Retest-Manifest = Test-retest reliability of strength scale score (3 wks. apart); Retest-Latent = Correlation of latent strength variables between test and retest (3 wks. apart) [estimates > 1 were limited to 1.00 for computation of means]. The row with means represents simple mathematical averages (without Fisher's r-to-Z' transformation).

Discussion

Overall, these results suggest that the IPIP-VIA-R short scales achieved sufficient levels of reliability in the present samples. Especially in view of the brevity of the scales and the fact that we selected content-valid IPIP items while preserving the substantive breadth of each character strength, many reliabilities were highly satisfactory. This applied to omega (the most appropriate reliability coefficient for these scales) and test-retest reliabilities.

Regarding the (less suitable) alpha reliabilities, let us compare IPIP-VIA-R scales to BFI-2 facet scales (Soto & John, 2017a, 2017b), which comprise four items with balanced keying for each of three domain-specific Big Five facets (see RQ6). In both countries, we observed somewhat lower values than reported for the BFI-2 facet scales of a US-dominated Internet validation sample, range = .59–.83 (*M* = .74; Soto & John, 2017a). Comparable α values resulted for the BFI-2 facet scales in a German Internet sample, range = .56–.84 (*M* = .73; Danner et al., 2019). Evidently, compared to IPIP-VIA-R, the BFI-2 scales meet alpha's assumptions better, because the items have rather narrow wordings and constitute rather homogeneous or factor-pure scales. By contrast, when looking at ω, the IPIP-VIA-R scales played in the same league as the German BFI-2 facet scales, range = .66–.87 (*M* = .78; Danner et al., 2019). Again, when looking at the retest-reliability coefficients, which do not assume homogeneous items in the scales, we obtained values (predominantly in Germany) that are comparable to those reported for the German BFI-2 facets, range = .58–.85 (*M* = .78; Danner et al., 2019).

Whereas most past reliability estimates for character strength scales cannot be trusted because their assumptions have not been tested, the internal consistency (ω) of IPIP-VIA-R short scales was satisfactory in the UK and in Germany, and test-retest reliabilities of scale scores (r_{tt}) confirmed this optimistic picture for Germany, while UK participants responded less temporally consistent. Not all scales performed equally well on all reliability estimates, though, and a few scales clearly had lower internal consistency (or test-retest reliability) than others. Especially for research in which the character strengths are used as predictors of life outcomes, the unreliability of scale scores can lead to attenuation bias, which requires statistical disattenuation. For such scenarios, we also recommend latent-variable models that control for method and error variance (see Lechner et al., 2021, for an overview).

Research Question 5:

Do IPIP-VIA-R Scales Function Equivalently Across Cultures?

Measurement invariance is a key prerequisite for research studying a construct of interest in different groups such as cultures—one simply cannot compare chopsticks with forks (Chen, 2008). Measurement invariance ensures that respondents with the same true standing on an underlying latent dimension have the same expected score on observed variables (Meredith, 1993). Measurement non-invariance introduces bias into cross-cultural comparisons of means, variances, or relations with predictors and outcomes.

Even though Peterson and Seligman (2004) selected character strengths for their VIA framework based on whether these strengths were universally acknowledged across different cultures, few studies have tested the measurement invariance of character strength scales across different cultures and languages. Some studies investigated the higher-order structure of character strengths in different cultures (e.g., McGrath, 2014, 2015, 2016; Ng et al., 2017) or structural relations among different strength scales (McGrath, 2016), yet these studies did not examine the cross-cultural invariance of each strength scale.

Personality and value inventories often show *limited* cross-cultural generalizability (e.g., expressions of spirituality; see MacDonald et al., 2015; Piedmont, 2007). Especially the comparability of latent means is often hard to achieve across cultures. Thus, it is important to establish whether our newly developed IPIP-VIA-R scales achieve levels of invariance that allow for meaningful comparisons across countries. Based on our measurement models developed in RQ3, we tested the configural, metric, and scalar measurement invariance of the 24 character strength scales across the UK and Germany (for details, see method). For well-constructed instruments, configural and metric invariance levels usually do not pose large challenges. The real challenge is scalar invariance. With some item adaptations affecting item difficulty (or vague quantifiers) to improve measurement quality in German items, we did not expect to find scalar invariance for all the scales.

Method

We tested measurement invariance by fitting 24 multi-group confirmatory factor analysis (MGCFA) models with robust maximum likelihood estimation (MLR). Establishing measurement invariance involves testing the fit of hierarchically nested, increasingly restricted measurement models across groups (Meredith, 1993; Vandenberg & Lance, 2000). If a specific level of *measurement invariance* holds, the next (higher) level can be inspected, until *invariance testing of structural parameters* becomes possible (Meredith, 1993). All aspects of the factor-analytic measurement model can be tested by introducing more and more equality constraints in MGCFA: Following Putnick and Bornstein (2016), the four measurement invariance steps that can be considered are: (1) *configural*, equivalence of model form (same item-factor configuration); (2) *metric* (weak factorial), equivalence of factor loadings (same units of measurement in scaling the construct); (3) *scalar* (strong factorial), equivalence of item intercepts or thresholds (absence of item difficulty bias); and (4) *residual* (strict or invariant uniqueness), equivalence of items' unique variances (amount of unsystematic error). When unfair measurement can be ruled out, one might proceed to testing structural invariance (factor means, variances, and covariances), if these are substantive questions of interest.

We limited our present analyses to the first three of these steps (configural, metric, and scalar) because scalar invariance is all that applied researchers typically seek. Scalar invariance ensures that constructs have the same meaning across cultures and allows for comparisons of latent variances, covariances, and means across cultures. If a specific level of measurement equivalence (e.g., scalar) does not hold, certain comparisons (e.g., of latent means) will not be permissible or at least be biased (Brown, 2015; Chen, 2008; Meredith, 1993; Steinmetz, 2013; Wu et al., 2007). In such cases, fitting a *partial* metric or scalar invariance model may still be possible: Even if some loadings and/or intercepts are unequal across groups, others may be invariant and one may achieve good model fit by lifting the equality constraints across groups of a few “problematic” (i.e., non-invariant) items (Byrne et al., 1989; Vandenberg & Lance, 2000).

To establish what level of invariance each scale could attain, we inspected 1) robust χ^2 statistics, scaled for deviation from normality (Satorra & Bentler, 2010); 2) fit-heuristics (CFI, RMSEA, SRMR; Bentler, 2007) considered less susceptible to sample size than χ^2 and $\Delta\chi^2$-tests, and 3) information criteria that favor models with a better accuracy-parsimony tradeoff (BIC; Schwarz, 1978). Instead of relying on a single criterion, we considered the whole profile of indices. We followed usual guidelines for model fit, which equal the cutoffs suggested for configural invariance (Hu & Bentler, 1999; Rutkowski & Svetina, 2014). We followed Chen’s (2007) criteria and accepted the metric level if the decrease in model fit in ΔCFI was not larger than .010 in combination with ΔRMSEA < .015 (or ΔSRMR < .030; cf. Cheung & Rensvold, 2002). Again, following Chen (2007), we accepted scalar equivalence if ΔCFI differences < .010 combined with ΔRMSEA < .015 (or ΔSRMR < .010). For the ultimate decision about accepting or rejecting an invariance level in ambiguous cases, we preferred information criteria over cutoff heuristics (see also Fan & Sivo, 2009; Marsh et al., 2004). When comparing invariance levels, lower BIC values indicate the better model when considering accuracy and parsimony jointly. Yet, ΔBIC $\leq$ 2 is hardly worth mentioning (i.e., models are effectively equivalent), ΔBIC between 2 and 5 is first evidence for one model being superior, but only a difference larger than 5 (or 10) is strong (or very strong) evidence that suggests discarding one of the models, that is, either the stricter or the more lenient invariance level (Raftery, 1995). For details on the approach, see SOM_IPIP-VIA-R_RQ5_Measurement-Invariance.pdf.

Results

Table 1.10 shows the outcomes of the testing procedure. BIC and fit heuristics mostly agreed. If not, then indices exceeded Chen's (2007) thresholds only marginally. If there was disagreement between BIC and fit heuristics, we accepted the higher invariance level (a) if lower BIC supported the more parsimonious model (which we indicate by presenting the accepted model in parentheses) or (b) if BIC increased only marginally, providing some (but not yet compelling) evidence against the more parsimonious (higher invariance) model (which we indicate by presenting the accepted model in double parentheses).

Configural Invariance. MGCFA models with the same configuration for all 24 scales (as established in RQ3) across countries fit well, confirming configural invariance. Yet, as already announced in RQ3, we first had to introduce one pair of correlated item residuals for three scales: Hope*, Leadership*, and Spirituality/Religiousness* (denoted by an asterisk in Table 1.10). Morizot et al. (2007) advised that residual correlations above $|r| > .20$ be flagged as evidence of a factor solution not adequately accounting for the data structure, thus signaling deviation from unidimensionality. The newly introduced residual correlations were meaningful and conveyed item overlap (in content or semantics; see Table 1.7): Hope items #1 and #4 reflect pessimism. Leadership items #1 and #2 reflect overlap regarding teamwork. Spirituality items #2 and #3 reflect meaning in life. After establishing fitting configural multigroup models, we proceeded to testing loading invariance.

Metric Invariance. Loadings were invariant for 23 out of 24 scales. The sole exception was Gratitude, which required one free loading to achieve *partial* metric invariance. Mplus modification indices (ModInd > 10) suggested relaxing the equality constraint on the first Gratitude item ("expressing thanks"). One can speculate here about cultural display rules that might govern "expressing thanks". After relaxing the equality constraint on the item, the (unstandardized) loading was 0.70 in the UK and 0.36 in Germany. Given that only one of the 96 items showed non-invariance, we judged metric invariance to be a given.

Scalar Invariance. Encouragingly, scalar invariance held for 11 out of 24 scales (see Table 1.10), despite minor BIC increases for two of these 11 scales. For the remaining 13 scales that did not attain full scalar invariance, we tested whether *partial* scalar invariance was attainable. With at least two invariant item intercepts, it is still possible to compare means of latent variables. We did this by freeing the equality constraint on the intercept with the highest modification index (usually when ModInd >10); if this did not yield an acceptable model, we freed a second item intercept. For seven out of 13 scales, partial scalar invariance resulted after freeing one intercept. The remaining six scales profited from a second free intercept, though for two scales—Self-Regulation and Citizenship/Teamwork—the decision to release a second parameter might be considered subjective: Compared to the models with one free intercept, BIC of the models with two free intercepts improved (by −2) in the presence of a moderate CFI increase (≤ +.016), while ModInd did not stick out (< 10).

Table 1.10

IPIP-VIA-R Measurement Invariance: UK vs Germany (Samples 4 and 5)

Model[a]	X^2 (*df*)	*p*	CFI	RMSEA	SRMR	BIC	ΔCFI	ΔRMSEA	ΔSRMR	ΔBIC	MI [b,c,d]
APP - configural	14.43 (10)	.154	.994	.031	.037	10977					YES
APP - metric	21.50 (13)	.064	.988	.037	.051	10964	−.006	.006	.014	−13	YES
APP - scalar	84.68 (16)	.000	.899	.095	.081	11010	−.089	.058	.030	46	NO
APP - (ν_2 free)	22.95 (15)	.085	.988	.033	.054	10952	.000	−.004	.003	−12	partial scalar
CAP - configural	37.92 (10)	.000	.959	.077	.050	11356					YES
CAP - metric	43.27 (13)	.000	.956	.070	.063	11342	−.003	−.007	.013	−14	YES
CAP - scalar	74.03 (16)	.000	.915	.087	.082	11352	−.041	.017	.019	10	NO
CAP - (ν_1, ν_3 free)	43.78 (14)	.000	.956	.067	.062	11336	.000	−.003	−.001	−6	partial scalar
CIT - configural	37.12 (10)	.000	.957	.076	.080	10625					YES
CIT - metric	42.12 (13)	.000	.954	.069	.092	10610	−.003	−.007	.012	−15	YES
CIT - scalar	77.91 (16)	.000	.902	.090	.102	10630	−.052	.021	.010	20	NO
CIT - (ν_1, ν_2 free)	42.86 (14)	.000	.954	.066	.095	10603	.000	−.003	.003	−7	partial scalar
CUR - configural	15.78 (10)	.106	.994	.035	.042	10490					YES
CUR - metric	20.08 (13)	.093	.992	.034	.055	10474	−.002	−.001	.013	−16	YES
CUR - scalar	70.52 (16)	.000	.939	.085	.066	10511	−.053	.051	.011	37	NO
CUR - (ν_1 free)	36.27 (15)	.002	.976	.055	.055	10478	−.016	.021	.000	4	((partial scalar))
EQU - configural	15.02 (10)	.131	.993	.033	.042	10401					YES
EQU - metric	31.21 (13)	.003	.974	.054	.070	10400	−.019	.021	.028	−1	(YES)
EQU - scalar	61.35 (16)	.000	.935	.077	.088	10416	−.039	.023	.018	16	NO
EQU - (ν_1 free)	37.50 (15)	.001	.986	.056	.073	10393	.012	.002	.003	−7	partial scalar
FOR - config	28.41 (10)	.002	.978	.062	.058	10567					YES
FOR - metric	38.28 (13)	.000	.970	.064	.071	10557	−.008	.002	.013	−10	YES
FOR - scalar	43.84 (16)	.000	.967	.061	.072	10542	−.003	−.003	.001	−15	scalar

(continues)

Model[a]	X^2 (*df*)	*p*	CFI	RMSEA	SRMR	BIC	ΔCFI	ΔRMSEA	ΔSRMR	ΔBIC	MI [b,c,d]
GRA - configural	34.29 (10)	.000	.969	.072	.074	10429					YES
GRA – metric	63.05 (13)	.000	.936	.090	.117	10441	−.033	.018	.043	12	NO
GRA - scalar	104.7 (16)	.000	.887	.108	.130	10465	−.049	.018	.013	24	NO
GRA - (λ_1 free)	42.09 (12)	.000	.962	.073	.083	10424	−.007	.001	.009	−5	partial metric
GRA - (λ_1, ν_2 free)	45.20 (14)	.000	.960	.068	.084	10413	−.002	−.005	.001	−17	partial scalar
HOP*- configural	24.82 (8)	.002	.983	.067	.046	10778					YES
HOP*- metric	33.58 (11)	.000	.977	.066	.057	10767	−.006	−.001	.011	−11	YES
HOP*- scalar	93.15 (14)	.000	.919	.109	.081	10804	−.058	.043	.024	37	NO
HOP*- (ν_1 free)	34.58 (13)	.001	.978	.059	.059	10754	.001	−.007	.002	−13	partial scalar
HUM - configural	15.66 (10)	.110	.994	.035	.031	10821					YES
HUM - metric	27.02 (13)	.012	.986	.048	.051	10812	−.008	.013	.020	−9	YES
HUM - scalar	51.79 (16)	.000	.963	.069	.062	10817	−.023	.021	.011	5	((scalar))
IND - configural	22.04 (10)	.015	.984	.050	.044	10341					YES
IND - metric	33.55 (13)	.001	.973	.058	.063	10334	−.011	.008	.019	−7	YES
IND - scalar	48.10 (16)	.000	.958	.065	.072	10332	−.015	.007	.009	−2	scalar
INT - configural	23.74 (10)	.008	.979	.054	.081	9665					YES
INT - metric	25.06 (13)	.023	.981	.044	.085	9645	.002	−.010	.004	−20	YES
INT - scalar	43.64 (16)	.000	.957	.060	.092	9650	−.024	.016	.007	5	((scalar))
JUD - configural	23.84 (10)	.008	.980	.054	.051	10506					YES
JUD - metric	23.46 (13)	.036	.985	.041	.052	10486	.005	−.013	.001	−20	YES
JUD - scalar	29.46 (16)	.021	.981	.042	.051	10472	−.004	.001	−.001	−14	scalar
KIN - configural	29.34 (10)	.001	.973	.064	.062	10700					YES
KIN - metric	33.65 (13)	.001	.971	.058	.071	10684	−.002	−.006	.009	−16	YES
KIN - scalar	68.91 (16)	.000	.926	.083	.085	10700	−.045	.025	.014	16	NO
KIN - (ν_1, ν_2 free)	35.42 (14)	.001	.970	.057	.077	10678	−.001	−.001	.006	−6	partial scalar

(continues)

Model[a]	X^2 (*df*)	*p*	CFI	RMSEA	SRMR	BIC	ΔCFI	ΔRMSEA	ΔSRMR	ΔBIC	MI [b,c,d]
LEA*- configural	14.82 (8)	.063	.992	.042	.037	10949					YES
LEA*- metric	19.66 (11)	.050	.990	.041	.040	10934	−.002	−.001	.003	−15	YES
LEA*- scalar	32.26 (14)	.004	.978	.052	.046	10927	−.012	.011	.006	−7	scalar
LOV - configural	19.06 (10)	.040	.991	.044	.041	10980					YES
LOV - metric	40.73 (13)	.000	.971	.067	.044	10982	−.020	.023	.003	2	(metric)
LOV - scalar	95.17 (16)	.000	.918	.102	.053	11016	−.053	.035	.009	34	NO
LOV - (ν_4, ν_2 free)	44.07 (14)	.000	.969	.067	.046	10978	−.002	.000	.002	−4	partial scalar
MOD - configural	40.45 (10)	.000	.944	.080	.050	11625					YES
MOD - metric	46.24 (13)	.000	.938	.073	.058	11611	−.006	−.007	.008	−14	YES
MOD - scalar	56.15 (16)	.000	.926	.073	.061	11599	−.012	.000	.003	−12	scalar
ORI - configural	33.27 (10)	.000	.975	.070	.044	10695					YES
ORI - metric	41.76 (13)	.000	.969	.068	.062	10685	−.006	−.002	.018	−10	YES
ORI - scalar	104.87 (16)	.000	.903	.108	.092	10728	−.066	.040	.030	43	NO
ORI - (ν_2, ν_4 free)	41.66 (14)	.000	.970	.064	.062	10678	.001	−.004	.000	−7	partial scalar
PER - configural	13.09 (10)	.218	.996	.026	.039	10237					YES
PER - metric	19.21 (13)	.117	.993	.032	.039	10224	−.003	.006	.000	−13	YES
PER - scalar	35.57 (16)	.003	.978	.051	.039	10221	−.015	.019	.000	−3	(scalar)
PRU - configural	32.37 (10)	.000	.970	.069	.056	10861					YES
PRU - metric	33.52 (13)	.001	.972	.058	.056	10842	.002	−.011	.000	−19	YES
PRU - scalar	94.39 (16)	.000	.893	.102	.081	10885	−.079	.044	.025	43	NO
PRU - (ν_2 free)	41.02 (15)	.000	.965	.060	.059	10837	−.007	.002	.003	−5	partial scalar
SEL - configural	21.86 (10)	.016	.982	.050	.042	11066					YES
SEL - metric	34.15 (13)	.001	.968	.059	.047	11059	−.014	.009	.005	−7	YES
SEL - scalar	102.8 (16)	.000	.869	.107	.070	11097	−.099	.048	.023	38	NO
SEL - (ν_3, ν_2 free)	37.76 (14)	.001	.964	.060	.052	11056	−.004	.001	.005	−3	partial scalar

(continues)

Model[a]	X^2 (*df*)	*p*	CFI	RMSEA	SRMR	BIC	ΔCFI	ΔRMSEA	ΔSRMR	ΔBIC	MI [b,c,d]
SOC - configural	22.35 (10)	.013	.985	.051	.040	10497					YES
SOC - metric	33.02 (13)	.002	.976	.057	.045	10488	−.009	.006	.005	−9	YES
SOC - scalar	35.58 (16)	.003	.977	.051	.048	10470	.001	−.006	.003	−18	scalar
SPI*- configural	12.42 (8)	.134	.995	.034	.032	12245					YES
SPI*- metric	19.26 (11)	.057	.990	.040	.039	12231	−.005	.006	.007	−14	YES
SPI*- scalar	37.58 (14)	.001	.971	.060	.048	12229	−.019	.020	.009	−2	(scalar)
VAL - configural	24.16 (10)	.007	.980	.055	.055	10729					YES
VAL - metric	29.75 (13)	.005	.977	.052	.048	10715	−.003	−.003	−.007	−14	YES
VAL - scalar	67.47 (16)	.000	.928	.082	.081	10733	−.049	.030	.033	18	NO
VAL - (ν_1 free)	38.96 (15)	.001	.967	.058	.064	10710	−.010	.006	.015	−5	partial scalar
ZES - configural	27.63 (10)	.002	.982	.061	.053	10735					YES
ZES - metric	31.34 (13)	.003	.982	.055	.055	10718	.000	−.006	.002	−17	YES
ZES - scalar	36.17 (16)	.003	.980	.052	.051	10702	−.002	−.003	−.004	−16	scalar

Note. [a] Greek letters reflect freed parameters in partial invariance models (e.g., λ_1/ ν_2 refer to free loading/intercept, respectively, of the indexed items #1 and #2); models with asterisk contain one pair of correlated residuals: HOP*, LEA*, SPI*; [b] MI = Measurement invariance level attainment: Configural invariance achieved for all scales; [c] Metric invariance achieved for all scales except: GRA (EQU & LOV permissible according to BIC, but set in parentheses), partial metric invariance achieved for GRA; [d] Scalar invariance achieved for eleven scales: FOR, JUD, LEA*, MOD, IND, PER, SOC, SPI*, ZES (HUM & INT tentatively permissible according to BIC, but set in double parentheses), partial scalar invariance achieved for APP, CAP, CIT, EQU, GRA, HOP, KIN, LOV, ORI, PRU, SEL, VAL (CUR tentatively permissible according to BIC, but set in double parentheses). Delta-fit values for partial invariance models are derived against the last accepted (full or partial) invariance model.

Discussion

All 24 IPIP-VIA-R scales achieved at least partial scalar invariance between Germany and the UK. This high degree of comparability across the two countries and languages is remarkable and shows that the efforts to ensure comparability during both the item selection and the translation and adaptation process were successful. Full scalar invariance implies that cross-country comparisons of both latent covariances and latent means are permissible (Bluemke et al., 2016), which applied to 11 IPIP-VIA-R scales. Partial scalar invariance implies that cross-country comparisons of latent covariances are permissible and—within limitations—also cross-country comparisons of latent means (Byrne et al., 1989). This applied to 13 IPIP-VIA-R scales (six scales with one noninvariant intercept; seven scales with two noninvariant intercepts). If researchers opt for the use of manifest IPIP-VIA-R scale scores (e.g., to allow comparisons with previous studies), cross-country comparisons are, based on our findings, only permissible with minor reservations, as (partial) scalar invariance does not ensure equivalent residual variances across countries (Bluemke et al., 2016).

Overall, then, findings from measurement invariance testing are encouraging. They support the suitability of the IPIP-VIA-R scales for cross-national research that, perhaps, generalizes to other English- and German-speaking countries and beyond. Before it can be claimed that the 24 character strengths represent cultural universals, research should test measurement (and structural) invariance across a larger and more diverse set of cultures and languages.

Research Question 6:

Are IPIP-VIA-R Scales Construct-Valid? A Nomological Network Perspective

Having closely defined the *empirical* framework for how to measure the strengths, we evaluated the construct validity of IPIP-VIA-R character strengths by locating them in a comprehensive nomological network (Cronbach & Meehl, 1955) with Big Five personality traits and basic human values. Whereas the relevance of Big Five is undisputed, the fundamental value circle proposed by Schwartz (1992, 1994) is equally relevant for value-laden character strengths, though research on associations with VIA strengths is scarcer. For our comprehensive approach, we first introduce our hypotheses about VIA strength associations with fundamental personality domains before we elaborate on the structure of value orientations and their relevance for VIA strengths.

Relations Between Character Strengths and Personality Traits. The five-factor model of personality is one fundamental model of basic traits: Extraversion, Agreeableness, Conscientiousness, Neuroticism/Negative Emotionality (or reversely: Emotional Stability), and Openness/Open-Mindedness (Goldberg, 1990; John et al., 1988). There is ample evidence that the Big Five reflect enduring dispositions that have a biological basis and manifest themselves in behavioral patterns. Though distinct trait expressions may be observed across cultures, researchers established five factors that typically underlie personality descriptors in numerous languages. Some researchers prefer including a sixth factor, whose label may depend on the criteria for lexical studies (e.g., Ashton et al., 2004; Ashton & Lee, 2020; Saucier, 2009; for a more nuanced (and hierarchical) view of personality traits, see Saucier & Iurino, 2020). We hypothesized for each of the 24 strengths about their most relevant Big Five domain.

The Big Five domains allow a reliable orientation within the dimensional sphere of personality (at least in Western samples). Several facets underlie each factor (Costa & McCrae, 1992, 2008), and in the case of the cross-culturally validated BFI-2 inventory (see Methods), three facets represent each Big Five domain (Soto & John, 2017a). The facet level provides highly specific traits (see Table 1.2a). Measured with fewer items than the domains, they are typically less reliable. Yet, combining multiple facets in multivariate regression, they are a powerful tool for increasing the predictive accuracy beyond that gained from using general domain scores (see also Mõttus et al., 2020). We therefore attend to predictive models that allow a comparison of associations of strengths with the Big Five framework both at the domain and at the facet level.

As demonstrated by earlier work (McGrath, 2014; Noftle et al., 2011), strengths are related to, but not identical with, the Big Five domains. As we outlined earlier, VIA character strengths reside at a lower (i.e., more fine-grained) level of abstraction than global Big Five personality traits, and also at a lower level than the four higher-order value clusters in Schwartz's (1992) hierarchical model of basic human values. Instead, character strengths reside on a similar level of abstraction as personality facets (though character strengths appear to be slightly broader than personality facets; cf. internal consistency in RQ4), and some individual strengths may partly measure the same traits as some personality facets (McGrath et al., 2020). Consistent with a (non-cognitive) skill view, character strengths are amenable to change, much like skills develop over the lifespan (Denissen et al., 2019; Specht et al., 2011).

We hypothesized *a priori* about the strongest relationships to any of the Big Five domains for each character strength, thereby providing a *confirmatory* test of this part of the nomological network. Given the longstanding dearth of research on relationships between strengths and specific personality facets at the time of our study (but see the recent progress on that front, e.g., McGrath et al., 2020; Ruch et al., 2021) and because it exceeds the scope of this study, we did not make any predictions for facets, rendering this part of the analysis *exploratory* in nature. Note that Peterson and Seligman's (2004) Tables 3.7 and 3.10 on the correspondence between character strengths and the Big Five framework and the Schwartz values remained incomplete, as they did not suggest for every character strength its most likely associate. Therefore, two raters in our research team hypothesized for each character strength its *strongest* positive or negative association with one of the Big Five domains (and with the Schwartz values, as discussed below). Raters agreed for 19 out of 24 strengths; disagreement was resolved through discussion. The one exception was Spirituality/Religiousness, because of its blend that might compound "subjective spirituality" and “orthodox religiosity”: Spirituality tends to be associated with high Openness, and Religiousness with low Openness (see Piedmont, 2007). Hence, raters disagreed on a *dominant* Big Five associate, but accepted two domains as validity evidence (see SOM_IPIP-VIA-R_RQ6_Nomological-Net-Hypotheses.xlxs).

Character Strengths and Basic Human Values. According to the theory of basic human values, ten motivationally distinct basic values represent core values that are culturally “shared conceptions of what is good and desirable in the culture” (Schwartz, 1992, 1994, 2006, p. 139; see Table 1.2b). Schwartz deemed the ten values universal—Power, Achievement, Hedonism, Stimulation, Self-Direction, Universalism, Benevolence, Tradition, Conformity, Security (for definitions, see Table 1.2b). These values derive from three universal requirements of human beings: their needs as biological organisms, their needs for coordinated social interaction, and their survival and welfare needs as groups (Schwartz, 1992, 1994, 2003a; Schwartz & Bilsky, 1987, 1990).

The ordering of ten human values provides a theory-driven circumplex model, where related values form value clusters in proximity: (1) Self-Enhancement versus Self-Transcendence, and (2) Openness to Change versus Conservation (Schwartz, 1992, 1994, 2003a; Schwartz & Boehnke, 2004). As two clusters each form opposing pairs, they jointly span a two-dimensional circumplex with (orthogonal) value orientation dimensions as axes, and the difference scores between opposing higher-order value clusters (e.g., Openness–Conservation) yield two dimensions of orthogonal value orientations. Despite the possibility of using finer value distinctions (Schwartz et al., 2012), the ten values *suffice* to comprehensively represent the clusters and reliably provide the higher-order dimensions (Schwartz, 2003b).

Dominant cultural values may form the context for developing individual value preferences. Though individuals and cultures differ in their priorities, value measurement functions equivalently across societies (Schwartz, 1994, 2006), and the basic structure of the

Schwartz value circle replicates (Bilsky et al., 2010; Schwartz & Sagiv, 1995). These assets make the four independently measured value clusters and the derived two-dimensional value orientation space (the associated circumplex) suitable candidates for validating character strengths. Comparing strengths to value clusters will be informative, as one can expect people to appreciate the kind of values that correspond to the strengths they possess. Likewise, they will tend to possess the exact strengths corresponding to what their value preferences are.

Just like basic human values, the VIA character strengths are cross-culturally inspired (Schwartz, 1992, 1994) and clearly value-laden (e.g., Lavy & Benish-Weisman, 2021). As character strengths are positively valued, thus related to value-laden personality traits, we posit that they reside at the intersection of two independent personality-relevant frameworks: fundamental personality traits and human values. Consequently, we simultaneously shed light on IPIP-VIA-R's relationships with the Big Five framework and Schwartz's human value clusters and orientations.

Our predictions of dominant associations for the four Schwartz human value clusters were less cogent than for the Big Five domains though (immediate agreement among two raters occurred for slightly more than half of the character strengths). This can mainly be attributed to the fact that the nature of the value circle made it possible to predict a general pattern of inverted signs of correlations for opposite clusters—a positive correlation with Conservation (or Self-Enhancement) implies a negative correlation with Openness to Change (or Self-Transcendence). Overall, we expected that most strengths would likely be positively associated with Self-Transcendence rather than Self-Enhancement, as all character strengths share the feature of being *socially sanctioned* traits.[12] Furthermore, for the character strengths scales, one should expect lower correlations with values compared to traits.

Scale Relationships. Prior to analyzing the theoretically expected trait and human value associations, we address one question that has not been convincingly answered for character strengths on the basis of viable strength measures: How are the 24 strengths themselves related to each other? How redundant are they? Would scales be associated similarly across the two countries? We provide a first glimpse at the scale correlations. Scale overlap is related to the strength hierarchy, which exceeds the present scope but receives attention elsewhere (e.g., Partsch et al., 2022).

Method

Participants. Samples 4 and 5 had provided responses to the IPIP-VIA-R survey and to Big Five and human values items. Despite merely two missing data points in Germany's IPIP-VIA-R (Curiosity, Zest), in both countries Schwartz value items were ("technically") missing more frequently (yet less than 5%), due to a "don't know" option. This resulted in different case numbers in the (pairwise-present) bivariate correlation matrix. Validity correlations for higher-order value clusters required working with complete-cases only. A sensitivity check based on SEM-based pseudo-indicator models (PIM; Rose et al., 2019), which accommodate values missing at random in manifest indicators by Full Information Maximum Likelihood (FIML), ruled out biases in the more intuitive approach presented below.

Big Five Measurement: BFI-2-S. The BFI-2-S is a 30-item questionnaire available, among many other languages, in English (Soto & John, 2017b) and German (Danner et al., 2019). It measures the Big Five domains of Extraversion, Agreeableness, Emotional Stability (or Negative Emotionality/Neuroticism), Conscientiousness, and Openness (Open-Mindedness). To provide both bandwidth and fidelity, the BFI-2 framework provides for each domain three underlying facets (see Table 1.2a). The BFI-2-S derives from its parent, the 60-item BFI-2. From there it inherits 50% of the items while keeping the factor structure robust. The response options are 5-point Likert-type (1 = *Disagree strongly,* 2 = *Disagree a little,* 3 = *Neutral; no opinion,* 4 = *Agree a little,* 5 = *Agree strongly*). The BFI-2-S has reportedly good reliability and converges very well with the full BFI-2 at the domain level. Thus, we hypothesized about the Big Five at the domain level, operationalized as manifest scale means after properly recoding reverse-keyed items (http://www.colby.edu/psych/wp-content/uploads/sites/50/2013/08/bfi2-form.pdf).

For facet and domain scores, we used the authors' aggregation rules. Five domain scores were derived by averaging scores of the balanced sets of six items per domain (Danner et al., 2019; Soto & John, 2017a, 2017b). Simultaneously, each domain score represents the aggregate of three equally weighted, balanced-keyed facets.

Basic Human Values Measurement: HVS. We used the Human Value Scale (HVS) taken from the European Social Survey (ESS; Schwartz, 2003a, 2003b). The HVS is based on one of the most common measures of Schwartz values, the 40-item Portrait Values Questionnaire (PVQ; Schwartz et al., 2001). The HVS includes 21 PVQ items, with a few wordings revised to better cover the content of the ten basic values. Each value is represented by two items, except for Universalism (three items). The HVS is available in English and German (Schwartz et al., 2015). Items are short verbal portraits of different people that convey the importance of different values to them, such as "It is important to him to be rich. He wants to have a lot of money and expensive things" (Power). We used the gender-specific versions with different pronouns. For each item, respondents indicate how similar the person described in the item is to them on a 6-point rating scale (1 = *Very much like me*, 2 = *Like me*, 3 = *Somewhat like me*, 4 = *A little like me*, 5 = *Not like me*, 6 = *Not like me at all* (plus an alternative response category *Don't know*). According to the protocol and ESS coding rules for HVS scales (Schwartz, 2003a, 2003b), when analyzing scale correlations, the ratings must be corrected for individuals' mean ratings beforehand. Only then they reflect value priorities free from individual response-style differences, corrected for individual differences in the use of the exclusively positively-keyed response scales, and yield the circumplex with balanced values and polar opposites. Individually centered (ipsative) scores express *relative* value preferences, which is crucial for correlational analyses and to maintain the circumplex structure.

The 21-item ESS-questionnaire is economic in its focus on capturing four value clusters reliably (but not each basic value). We computed four indexes for Self-Enhancement, Self-Transcendence, Conservation, and Openness to Change. In line with ESS-prescriptions, higher-order dimension scores are formed from difference scores between value clusters, such that high scores reflect the relative preference of values of Self-Transcendence over Self-Enhancement, and Openness to Change over Conservation (Schwartz, 2013).

Analytical Approach. We computed Pearson correlations among the IPIP-VIA-R scale scores and between each IPIP-VIA-R scale score and both the Big Five domain scores and the HVS cluster scores, respectively. This allows for comparing our findings with previous ones of similarly analyzed, single- or multi-item strength scales (e.g., McGrath et al., 2020; Ruch et al., 2007, 2014). To explore the overlap between each strength and basic personality variables, we also computed regression models that predict each strength by the primary Big Five domain (complemented by all five domains), alternatively by the primary facet (complemented by all 15 facets). For a comparison of IPIP-VIA-R validity coefficients with McGrath et al. (2020), we disattenuated the short scale correlations with the Big Five framework also for unreliability (single and double disattenuation for predictor and criterion variable). Based on multiple-*R* values (and R^2 as variance explained) we identified which character strengths overlap (rather strongly or weakly) with the pool of all Big Five domains or all 15 facets, respectively. Finally, to further examine the relationship between strengths and values we used cluster difference scores representing the two higher-order value dimensions. Using the correlation coefficients between strengths and higher-order value dimensions as a basis for coordinates (ranging between −1 and +1) allows projecting each strength into the two-dimensional space of higher-order value orientations: Self-Transcendence versus Self-Enhancement (approximating a social growth focus) and Openness to Change versus Conservation (approximating a personal growth focus).

Results

Interrelationships Among Strengths. Character strength scales were almost exclusively positively correlated (see Table 1.11). That is, they formed a positive manifold. The mean of the average correlation for each IPIP-VIA-R scale with all other scales was r = .37 (.31) in the UK (Germany). According to each scale's average correlation, in both countries Modesty and Prudence had the lowest overlap with other scales, while the greatest overlap with other scales in both countries had Gratitude, Perspective, Curiosity. The highest correlation coefficients found for *any pairs* of scales in Germany replicated in the UK: Curiosity with Love of Learning (.62 and .58, in Germany and the UK, respectively), and Zest with Hope (.61 and .61). In the UK, though, other scales correlated slightly higher: Gratitude with Kindness (.65), and Leadership with Social Intelligence (.64), which also correlated substantially but somewhat weaker in Germany (.54 and .45, respectively). When scales were unrelated, such as Prudence with Spirituality/Religiousness, this held in both countries (.06); likewise, Self-Regulation with Humor (.16 and .12 in the UK and Germany, respectively).

These strength correlations were generally similar across countries. The average difference between correlation coefficients across countries amounted to 0.06 (SD = 0.07). There were only few strengths for which the differences were larger. For example, Forgiveness correlated significantly with Judgment and Spirituality/Religiousness at r = .31 and .38 in the UK, but in Germany Forgiveness did not correlate significantly with Judgment, r = .02, and much lower with Spirituality/Religiousness, r = .13. Leadership correlated significantly (and negatively so) with Modesty in Germany, but not so in the UK (r = −.19 vs .03).

Table 1.11

Associations of Character Strength Scales: IPIP-VIA-R Inter-Correlations (Samples 4 and 5)

	APP	CAP	CIT	CUR[a]	EQU	FOR	GRA	HOP	HUM	IND	INT	JUD	KIN	LEA	LOV	MOD	ORI	PER	PRU	SEL	SOC	SPI	VAL	ZES[a]
APP	–	.45	.34	.49	.41	.22	.52	.34	.39	.34	.43	.38	.40	.29	.46	.15	.42	.45	.22	.30	.41	.34	.24	.41
CAP	.42	–	.49	.42	.35	.34	.53	.39	.49	.35	.31	.30	.42	.30	.32	.11	.39	.40	.12	.18	.45	.24	.26	.55
CIT	.42	.57	–	.46	.39	.34	.45	.31	.45	.41	.41	.32	.52	.40	.32	.13	.32	.42	.13	.18	.39	.17	.31	.46
CUR[a]	.53	.49	.57	–	.32	.31	.47	.45	.51	.41	.34	.32	.51	.45	.62	−.10	.57	.52	−.02	.22	.44	.28	.31	.59
EQU	.44	.34	.47	.46	–	.36	.48	.34	.23	.39	.50	.28	.48	.17	.30	.32	.17	.30	.26	.33	.28	.24	.15	.37
FOR	.23	.36	.42	.37	.50	–	.39	.43	.23	.13	.17	.02	.41	.17	.24	.04	.13	.21	−.03	.17	.20	.13	.07	.44
GRA	.57	.56	.60	.53	.58	.41	–	.39	.44	.40	.48	.35	.54	.24	.40	.14	.32	.48	.20	.23	.41	.36	.24	.50
HOP	.27	.39	.44	.47	.33	.44	.38	–	.43	.41	.26	.20	.38	.40	.39	.03	.37	.47	−.14	.34	.41	.24	.28	.61
HUM	.40	.46	.56	.55	.42	.32	.50	.44	–	.38	.28	.33	.46	.50	.43	−.05	.57	.51	−.14	.12	.52	.19	.42	.57
IND	.32	.39	.48	.48	.42	.32	.47	.43	.41	–	.50	.43	.38	.38	.41	.18	.35	.51	.20	.37	.35	.11	.44	.49
INT	.46	.33	.52	.41	.54	.30	.60	.29	.40	.55	–	.42	.47	.18	.30	.35	.23	.40	.31	.35	.27	.16	.35	.31
JUD	.48	.33	.45	.50	.45	.31	.51	.31	.41	.48	.51	–	.34	.34	.37	.08	.37	.50	.30	.19	.48	.16	.34	.24
KIN	.45	.47	.61	.54	.60	.46	.65	.42	.48	.48	.58	.48	–	.38	.40	.22	.36	.42	.10	.23	.45	.27	.33	.46
LEA	.31	.43	.56	.48	.23	.28	.38	.45	.50	.45	.33	.41	.40	–	.44	−.19	.51	.49	−.16	.16	.45	.20	.39	.41
LOV	.48	.33	.46	.58	.50	.34	.48	.25	.42	.39	.40	.52	.44	.35	–	−.04	.46	.44	.02	.17	.38	.20	.28	.44
MOD	.25	.09	.19	.04	.27	.06	.34	.02	.14	.23	.41	.25	.28	.03	.12	–	−.16	−.03	.34	.18	−.08	.02	−.04	.05
ORI	.40	.36	.43	.61	.33	.29	.35	.37	.43	.41	.33	.48	.37	.51	.51	−.12	–	.54	−.14	.15	.49	.22	.37	.38
PER	.44	.46	.49	.56	.41	.30	.58	.43	.48	.51	.51	.62	.53	.55	.45	.17	.49	–	.07	.32	.55	.25	.37	.49
PRU	.28	.10	.14	.07	.26	.04	.24	−.09	.07	.20	.35	.37	.19	.01	.12	.34	−.03	.24	–	.19	.04	.06	−.07	−.04
SEL	.20	.13	.21	.26	.24	.31	.23	.32	.16	.37	.40	.34	.24	.23	.17	.21	.23	.34	.31	–	.24	.15	.12	.28
SOC	.44	.55	.57	.57	.40	.37	.52	.51	.57	.49	.39	.57	.51	.64	.41	.07	.51	.61	.11	.28	–	.27	.32	.42
SPI	.25	.35	.36	.35	.28	.38	.33	.36	.28	.31	.26	.19	.30	.27	.27	.01	.28	.27	.06	.22	.39	–	.05	.29
VAL	.26	.30	.31	.39	.27	.19	.33	.32	.38	.36	.27	.38	.27	.45	.35	−.10	.51	.47	−.04	.10	.41	.19	–	.24
ZES[a]	.28	.45	.47	.56	.35	.41	.39	.61	.46	.48	.33	.32	.41	.43	.32	−.03	.40	.39	.02	.37	.51	.42	.27	–

Note: N = 950 (474 in UK; 476 in DE); [a] N = 475 (in DE); UK = below diagonal; DE = above diagonal. For UK and DE, $p < .05$ if $|r| > .09$; $p < .01$ if $|r| > .12$, $p < .001$ if $|r| > .16$.

Relationships With Big Five Domains. As evident from Table 1.12, most character strengths were clearly linked to the (primary) Big Five domain as expected. For 14 scales, both countries confirmed the respective relationship as the strongest one. The strength of associations between IPIP-VIA-R scales and all BFI-2-S domain scores usually fell in the mid-range (.30–.60). Of the many correlations, let us highlight the strongest (absolute) correlation coefficients found between each domain and any strength (for the UK/Germany, respectively): Leadership with Extraversion (.60/.66), Kindness with Agreeableness (.64/.61), Industry/Perseverance with Conscientiousness (.59/.69), Hope with Negative Emotionality/Emotional Stability (58./.73.; alternatively Zest: .58/.69), and Originality with Open-Mindedness (.62/.67). These scale correlation coefficients mark the upper boundary of convergent validity coefficients for manifest IPIP-VIA-R strength scales regarding the Big Five domains. All the other strengths also showed substantial convergence with Big Five domains but, at the same time, the coefficients revealed that most strengths cannot just reflect a single (or multiple) Big Five domain(s) in disguise (square, for instance, the adjusted multiple-*R* values in Table 1.13 that result from multiple regression models predicting each IPIP-VIA-R scale with all Big Five domains discussed in the next subsection, which hardly cross the threshold of 50% of explained variance). On average, the absolute scale intercorrelation coefficients amounted to $\bar{r}$ = .51 (*SD* = 0.09) and = .52 (*SD* = 0.12), for the UK and Germany, respectively. For more fine-grained insights, please refer to Table 1.12.

Our presentation so far highlights the convergent validity for most IPIP-VIA-R strength scales for fundamental personality traits. In those cases where our hypotheses did not fully bear out, that is, where a different domain produced a higher association than the predicted one, the win was usually by a small margin only. For instance, the correlation between Social Intelligence and Agreeableness emerged as envisioned (.45/.39), but Extraversion correlated marginally higher in both countries (.48/.43). The unexpected relationship is plausible in hindsight, given the high level of Sociability found in extraverts.

As for another example, Modesty correlated with Extraversion as expected (−.30 in both countries), yet a similar relationship with Agreeableness surfaced in the UK (.36) and in Germany (.30). (Note that for these strengths Peterson and Seligman (2004) did not forecast any associations in their Table 3.7 either.) In the case of Humor, the two countries disagreed on the primary domain (.43 for Agreeableness in the UK; .58 for Extraversion in Germany). The latter correlation for Germany deviates from our own expectation but confirms Peterson and Seligman's (2004; Table 3.7). Rather than interpreting the pattern as outright disconfirmation or considering such scales as "weak", these associations unveil some difficulty in making clear-cut predictions about primary domains. The one scale we had anticipated to behave unruly in this regard, Spirituality/Religiousness, indeed correlated low and on par with all domains (−.17–.36; cf. McGrath et al., 2020).

Two "real" disconfirmations question our understanding of Equity and Valor/Bravery. Our intuition was to associate Equity primarily with Conscientiousness (.35/.41), as we assumed a motivation to act in a responsible manner in social contexts as the motor behind this association. Yet, both countries agreed that Agreeableness (.63/.61) was the primary associate, thereby highlighting the inner drive to maintain social peace associated with Equity. Similarly, Valor/Bravery appears to require much less the absence of Anxiety or Negative Emotionality (−.24/−.28) than the expression of Extraversion (.37/.41), probably as Assertiveness in extraverts helps challenge others and prevail over opponents in social encounters.

Nearly all strengths correlated with *all* Big Five domains. Rarely did a strength *not* correlate with one of the domains (e.g., Integrity was largely unrelated to Extraversion). The presence of multiple domain associations yields first evidence that strengths reside at intersections of basic traits and cut right through the variable space of personality, though character strengths and the lexical personality space are not redundant (as we will see next).

Table 1.12

IPIP-VIA-R Nomological Net: Cross-Country Comparison of Correlations of Character Strengths with Big Five Domains and Human Value Clusters (Samples 4 and 5)

	E		A		C		N		O		S-T		S-E		O-C		CON	
Scale	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE
APP	.14	.25	.40	.41	.31	.35	−.12	−.29	**.51**	**.48**	**.38**	**.34**	−.31	−.26	.03	.00	−.06	−.08
CAP	.36	.35	**.47**	**.45**	.35	.35	−.31	−.38	.31	.29	**.25**	**.24**	−.22	−.23	.17	.08	−.18	−.10
CIT	.38	.36	**.54**	**.45**	.45	.33	−.31	−.37	.37	.25	**.35**	**.29**	−.27	−.16	.08	.01	−.12	−.12
CUR	.43	.50[a]	.42	.41[a]	.36	.33[a]	−.38	−.45[a]	**.52**	**.57[a]**	.23	.22[c]	−.24	−.09[b]	**.32**	.20[c]	−.27	**−.31[d]**
EQU	.12	.12	**.63**	**.61**	.35	.41	−.19	−.31	.41	.24	**.60**	**.50**	−.44	−.39	−.03	−.14	−.11	.06
FOR	.23	.21	**.57**	**.42**	.34	.13	−.34	−.41	.34	.15	**.38**	**.23**	−.29	−.14	.05	.03	−.11	−.11
GRA	.17	.24	**.61**	**.54**	.42	.37	−.24	−.33	.36	.32	**.48**	**.35**	−.43	−.28	.00	−.04	−.02	−.03
HOP	.46	.43	.44	.38	.37	.32	**−.58**	**−.73**	.30	.39	.17	.17	−.24	−.16	**.27**	**.19**	−.17	**−.19**
HUM	.40	**.58**	**.43**	.38	.34	.25	−.34	−.46	.35	.45	.25	.10	**−.27**	−.09	**.27**	.31	−.21	**−.32**
IND	.27	.38	.45	.32	**.59**	**.69**	−.37	−.53	.31	.32	.21	**.13**	**−.28**	**−.13**	.07	.00	−.01	−.01
INT	.07	.16	**.56**	.48	.54	**.55**	−.25	−.28	.28	.21	.40	.37	**−.46**	**−.39**	−.09	−.07	.12	.08
JUD	.14	.19	.43	.26	**.48**	**.37**	−.26	−.23	.43	.34	.29	**.24**	**−.32**	−.13	.00	−.13	.02	.02
KIN	.20	.32	**.64**	**.61**	.43	.38	−.24	−.39	.36	.35	**.50**	**.46**	−.46	−.38	.06	.03	−.08	−.09
LEA	**.60**	**.66**	.30	.22	.39	.31	−.42	−.45	.34	.46	.07	.03	−.08	.08	**.23**	.21	−.18	**−.33**
LOV	.19	.41	.39	.34	.31	.27	−.10	−.35	**.52**	**.56**	**.36**	.21	−.25	−.02	.12	.11	−.20	**−.28**
MOD	−.30	**−.30**	**.36**	**.30**	.29	.25	.00	−.07	−.02	−.22	.37	.31	**−.44**	**−.52**	−.36	−.24	.38	.43
ORI	.42	.52	.27	.24	.28	.24	−.27	−.32	**.62**	**.67**	.11	.06	−.06	.00	**.38**	.33	**−.38**	**−.37**
PER	.33	.46	.43	.38	**.48**	.45	−.41	**−.52**	.39	**.52**	.26	**.20**	**−.31**	−.07	.11	.07	−.04	−.18

(continues)

	E		A		C		N		O		S-T		S-E		O-C		CON	
Scale	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE	UK	DE
PRU	−.25	−.30	.23	.21	**.33**	**.33**	.02	.11	.05	−.12	.17	.21	−.26	−.20	−.42	**−.52**	**.45**	.51
SEL	.17	.15	.34	.27	**.52**	**.43**	−.43	−.38	.17	.17	.16	.20	**−.30**	**−.26**	−.06	−.07	.18	.12
SOC	**.48**	.43	.45	.39	.47	.32	−.39	−.36	.40	**.49**	**.21**	**.25**	−.24	−.11	.20	.13	−.14	**−.25**
SPI	.25	.21	**.36**	**.29**	.21	.19	−.25	−.17	.25	**.29**	.20	**.20**	**−.26**	−.16	.04	−.11	.05	.06
VAL	**.37**	**.41**	.15	.11	.20	.27	−.24	−.28	.36	.32	.18	.14	−.13	−.12	.28	.23	**−.29**	**−.26**
ZES	.56	.54[a]	.47	.48[a]	.47	.44[a]	**−.58**	**−.69[a]**	.33	.37[a]	.11	.17[c]	−.19	**−.18[b]**	**.30**	**.18[c]**	−.19	−.17[d]
N	474	476	474	476	474	476	474	476	474	476	438	440	444	448	442	440	423	431

Note. [a] N = 475, [b] N = 447, [c] N = 439, [d] N = 430; UK = United Kingdom, DE = Germany; Scale = IPIP-VIA-R strength; Big Five Domains: E = Extraversion, A = Agreeableness, C = Conscientiousness, N = Neg. Emotionality (Neuroticism), O = Open-Mindedness; Schwartz Value Higher-Order Dimensions (centered): S-T = Self-Transcendence, S-E= Self-Enhancement, O-C = Openness to Change, CON = Conservation, with Schwartz value cluster means based on ipsatively centered item scores. For all coefficients, $p < .05$ if $|r| > .09$; $p < .01$ if $|r| > .12$, $p < .001$ if $|r| > .16$, except $r_{\text{PRU~O|DE}}$ and $r_{\text{PRU~S-T|UK}}$ ($p < .01$). Underlining refers to the hypothesized strongest (absolute) correlation per country with respect to two frameworks (Big Five and Schwartz human values); bold type font reflects the empirically highest coefficients per country within each framework; underlined bold font indicates a confirmed hypothesis (accuracy of comparisons: only to the level of digits as shown).

Conceptual Overlap With Big Five Domains and Facets. To shed light on the redundancy between strengths and basic personality variables, in line with McGrath et al.'s (2020) approach, we regressed each strength on (a) the primary domain, (b) all domains, (c) the primary facet, and (d) all facets. Looking at the Big Five facet level allows for a more fine-grained location of the strengths in the personality space. The facet-level analysis is informative, because Big Five facets and character strengths are similarly abstract and allow for rather symmetrical abstractness of predictor and criterion variables. Also, using 15 facet scores rather than five domain scores as predictors in regression models yields more statistical power for explaining the variance in strength scales. We additionally corrected correlations for unreliability of the predictor (single disattenuation) and, on top, for the unreliability of the criterion (double disattenuation; see also SOM_IPIP-VIA-R-RQ6_Validity-Correlations-BFI-2-S.xlsx).

Tables 1.13a and 1.13b show in descending order, separately for each country, which domains and facets overlapped mostly with the strengths. To summarize the picture: At the level of domains, the countries converged on the single domain with the largest correlation for 19 strengths. This number increased to 23 when including the domain with the second largest correlation from Table 1.12. The one exceptional scale was Perspective, which showed mid-sized correlations regardless of the domain, making the pattern volatile. At the level of facets, the country-specific analyses converged for 15 strengths on the single facet with the largest correlation, which increased to 18 when counting other facets from the same domain as cross-validated. For instance, Hope correlated mostly with Depression in the UK and with Anxiety in Germany, but both Big Five facets also belong to the same domain, Negative Emotionality.

Although rare, it could happen that the facet with the strongest correlation with a strength did not come from the domain which produced the strongest correlation. For instance, Appreciation for Beauty and Excellence, which was related primarily to Open-Mindedness, was best predicted in Germany by the Agreeableness-facet Compassion.

Conversely, the facet Creative Imagination predicted Humor best in both countries, possibly due to a creative playfulness aspect, though Humor's primary domain was not Open-Mindedness, but Agreeableness in the UK and Extraversion in Germany.

Interestingly, each strength's strongest facet correlation tended to be of similar size as the correlation with its primary domain. For most strengths slightly higher coefficients resulted for the primary facet (r_F) compared to the primary domain (r_D) (e.g., r_F = .66 versus r_D = .62 and r_F = .75 versus r_D = .67 for Originality in the UK and Germany, respectively), while for few strengths the opposite held (e.g., r_F = .56 versus r_D = .63 and r_F = .57 versus r_D = .61 for Equity). This finding suggests that, from the perspective of each strength, the primary relationship with a Big Five domain is mostly driven by a specific facet that can be identified: For instance, Originality is not just related to Open-Mindedness broadly, but specifically reflects the facet Creative Imagination. This pattern conforms to strengths residing at a lower level than Big Five domains in the nomological network, while being roughly as abstract as personality facets.

As each personality facet was measured with two items, hence with less precision than each domain based on six items, we observed overcorrection for some disattenuated facet correlations that might be mistaken as perfect conceptual overlap. Note that the *adjusted-R* coefficients from multiple regression are more informative and contradict the notion that strengths are identical with personality facets. Notably, with both IPIP-VIA-R and BFI-2-S controlling for ARS at the scale level, none of the strengths could be completely forecast by taking in all Big Five domains (or facets). Squaring the reported multiple-*R* correlation coefficients from Tables 1.13a and 1.13b shows that all Big Five domains (or facets) together hardly ever explained more than 50% of the variance in strength scales. The comparison of adjusted multiple-*R* coefficients to zero-order correlations further demonstrates that adding several domains (or multiple facets) hardly improved the statistical predictions (despite facet-level models being statistically powerful).

Table 1.13a

IPIP-VIA-R Construct Validity: Multiple Regression on Big Five Domains and Facets (Sample 4: UK)

Big Five Domains						Big Five Facets					
Scale	Domain	r	$r^{\dagger}$	$r^{\ddagger}$	R_{adj}	Scale	Facet	r	$r^{\dagger}$	$r^{\ddagger}$	R_{adj}
KIN	A	.64	.75	.91	.67	ORI	O-c	.66	.87	1.04•	.71
EQU	A	.63	.74	.88	.67	IND	C-p	.64	.87	1.04•	.68
ORI	O	.62	.73	.87	.67	ZES	N-d	.61	.74	.88	.77
GRA	A	.61	.72	.84	.64	INT	C-r	.60	1.02•	1.19•	.67
LEA	E	.60	.72	.87	.66	GRA	A-c	.58	.87	1.02•	.68
IND	C	.59	.68	.81	.63	EQU	A-r	.56	.83	.97	.67
ZES	N	.58	.63	.75	.73	LEA	E-a	.56	.75	.90	.71
HOP	N	.58	.62	.74	.67	KIN	A-c	.55	.83	1.00•	.68
FOR	A	.57	.67	.81	.61	HOP	N-d	.55	.67	.79	.69
INT	A	.56	.65	.76	.64	FOR	A-t	.54	.78	.95	.62
CIT	A	.54	.64	.79	.65	JUD	C-r	.50	.85	1.05•	.60
CUR	O	.52	.62	.75	.65	CUR	O-c	.49	.64	.79	.67
LOV	O	.52	.61	.75	.58	CIT	A-c	.48	.72	.89	.67
SEL	C	.52	.59	.76	.56	PER	C-r	.48	.81	.98	.64
APP	O	.51	.60	.75	.57	CAP	A-c	.47	.71	.86	.64
SOC	E	.48	.58	.68	.66	SOC	C-r	.45	.77	.91	.69
JUD	C	.48	.55	.68	.59	APP	O-i	.45	.72	.90	.59
PER	C	.48	.55	.66	.60	LOV	O-i	.44	.71	.87	.59
CAP	A	.47	.55	.67	.56	MOD	A-r	.43	.64	.82	.57
HUM	A	.43	.50	.59	.56	SEL	N-e	.43	.49	.64	.59
VAL	E	.37	.45	.56	.45	VAL	E-a	.41	.54	.67	.54
SPI	A	.36	.43	.55	.42	HUM	O-c	.40	.53	.63	.59
MOD	A	.36	.42	.54	.54	PRU	C-r	.37	.63	.84	.54
PRU	C	.33	.38	.51	.49	SPI	A-c	.32	.49	.62	.45
Mean		**.51**	**.60**	**.73**	**.60**	***Mean***		**.50**	**.73**	**.87**	**.64**

Notes. N = 474. Scale = IPIP-VIA-R strength; r = max. (uncorrected) bivariate absolute correlation with BFI-2-S domain or facet (in descending order); $r^{\dagger}$ = single disattenuation for unreliability in BFI-2-S domain (Cronbach's Alpha for six items) or BFI-2-S facet (Spearman-Brown-correction for two-item correlation; see Eisinga et al., 2013); $r^{\ddagger}$ = double disattenuation for unreliability in BFI-2-S and strength scale (Omega estimate). R_{adj} = multiple-R (square root of R^2_{adj}) for full models regressing strength on five BFI-2-S domains or 15 BFI-2-S facets); • overadjustment (estimates > 1 were limited to 1.00 for computation of means).

Table 1.13b

IPIP-VIA-R Construct Validity: Multiple-Regression on Big Five Domains and Facets (Sample 5: DE)

Big Five Domains						Big Five Facets					
Scale	Domain	r	$r^{\dagger}$	$r^{\ddagger}$	R_{adj}	Scale	Facet	r	$r^{\dagger}$	$r^{\ddagger}$	R_{adj}
HOP	N	.73	.80	.94	.75	ORI	O-c	.75	.91	1.05•	.77
ZES [a]	N	.69	.76	.88	.76	IND	C-p	.70	.93	1.06•	.77
IND	C	.69	.77	.89	.74	ZES [a]	N-d	.69	.82	.94	.80
ORI	O	.67	.78	.90	.70	HOP	N-a	.67	.94	1.10•	.76
LEA	E	.66	.78	.93	.69	LEA	E-a	.63	.77	.92	.71
EQU	A	.61	.73	.87	.64	INT	C-r	.61	.87	1.05•	.65
KIN	A	.61	.72	.93	.65	KIN	A-c	.60	.92	1.18•	.67
HUM	E	.58	.68	.78	.65	EQU	A-r	.57	.80	.96	.65
CUR [a]	O	.57	.66	.76	.67	CUR [a]	E-e	.56	.74	.84	.69
LOV	O	.56	.66	.80	.61	PER	O-c	.54	.66	.80	.69
INT	C	.55	.62	.74	.61	GRA	A-c	.53	.82	1.00•	.62
GRA	A	.54	.64	.79	.58	HUM	O-c	.53	.64	.72	.69
PER	N	.52	.57	.70	.67	CAP	A-c	.48	.74	.91	.61
SOC	O	.49	.58	.73	.59	CIT	A-c	.48	.73	.95	.56
APP	O	.48	.56	.71	.58	SOC	A-c	.47	.72	.91	.62
CAP	A	.45	.54	.66	.53	LOV	O-c	.46	.56	.68	.62
CIT	A	.45	.53	.69	.52	APP	A-c	.42	.65	.82	.61
SEL	C	.43	.48	.66	.47	SEL	C-p	.40	.53	.74	.51
FOR	A	.42	.50	.57	.52	JUD	C-r	.40	.57	.76	.53
VAL	E	.41	.48	.58	.45	FOR	A-t	.40	.75	.86	.53
JUD	C	.37	.41	.54	.45	MOD	E-a	.36	.45	.59	.57
PRU	C	.33	.37	.44	.56	VAL	E-a	.36	.44	.53	.46
MOD	E	.30	.36	.47	.55	PRU	C-o	.33	.38	.45	.61
SPI	A	.29	.35	.40	.36	SPI	A-c	.32	.49	.57	.37
Mean		**.52**	**.60**	**.72**	**.60**	***Mean***		**.51**	**.70**	**.83**	**.63**

Notes. N = 476; [a] N = 475. Scale = IPIP-VIA-R strength scale; r = max. (uncorrected) absolute bivariate correlation with BFI-2-S domain or facet (in descending order); $r^{\dagger}$ = single disattenuation for unreliability in BFI-2-S domain (Cronbach's Alpha for six items) or BFI-2-S facet (Spearman-Brown-correction for two-item correlation; see Eisinga et al., 2013); $r^{\ddagger}$ = double disattenuation for unreliability in BFI-2-S and strength scale (using omega estimate for the latter). R_{adj} = multiple-R (square root of R^2_{adj}) when regressing strength on five BFI-2-S domains or 15 BFI-2-S facets); • overadjustment (estimates > 1 were limited to 1.00 for computation of means).

Some character strengths *did* overlap notably with single personality facets, but only four (seven) primary facet correlations exceeded uncorrected values of .60 (i.e., McGrath et al.'s (2020) cutoff for "likely redundancy") in the UK (Germany). However, statistical cutoffs alone can never demarcate if concepts are close to being identical, not even after correction for scale unreliability. Take Love of Learning with its disattenuated correlation of .87 with the Big Five facet Intellectual Curiosity in the UK as an example. One can easily envision that Intellectual Curiosity—comprising the BFI-2-S items "being a complex, deep thinker" and "having little interest in abstract ideas"—is associated with Love of Learning, but from a conceptional point of view the personality facet and strength zoom in on different cognitive motivations and processes (i.e., focus on thinking versus learning). And yet, we also observed both empirical and conceptual closeness between other scales, such as between Compassion—measured with the BFI-2-S items "being compassionate, having a soft heart" and "can be cold and uncaring"—and Kindness as a strength. Such concepts may indeed be exchangeable. They require close inspections of scale content before preferring either of the scales.

Relationships With Human Value Dimensions. As regards basic human values, the HVS scales showed the typical pattern of reversed correlations across those value clusters that oppose each other and form the ends of the higher-order value dimensions (see Table 1.12 and Figure 1.3). For Self-Transcendence versus Self-Enhancement, we found the following pattern: Almost without exception, character strengths correlated positively with Self-Transcendence and negatively with Self-Enhancement. Accordingly, when using the respective difference scores (i.e., Self-Transcendence minus Self-Enhancement), character strengths predominantly strove towards the positive pole, with Equity, Kindness, and Modesty being among the strongest positive (negative) correlates of Self-Transcendence (Self-Enhancement) and the most extremely/outwards located strengths, while Leadership and Originality being among the weakest correlating strengths located close to the origin. Notably, only Leadership in Germany slightly strove towards the negative (i.e., Self-Enhancement) pole (see Figure 1.3b). This pattern supports the social value recognized in, and ascribed to, character strengths, even of those strengths that are rather intellectual in nature (e.g., Love of Learning).

For Openness to Change versus Conservation, we found the following pattern: While some strengths correlated positively with Openness to Change and negatively with Conservation (e.g., Curiosity, Humor, Originality, Valor/Bravery), others correlated positively with Conservation and negatively with Openness to Change (e.g., Modesty, Prudence, Self-Regulation), leading to a rather even spread of strengths along the horizontal axis in Figure 1.3 with the former (latter) strengths striving towards the positive (negative) pole on the second higher-order dimension (i.e., Openness to Change minus Conservation). While some strengths showed *strong* positive correlations with Openness to Change and negative correlations with Conservation (e.g., Originality) or vice versa (e.g., Prudence), many character strengths showed comparably *weak* correlations with the both opposing clusters (e.g., Gratitude in both countries). Accordingly, in Figure 1.3, Originality strove towards the Openness pole, Prudence strove towards the Conservation pole, and Gratitude was located at the origin on this higher-order dimension. To conclude, most strengths manifested their potential through transcending rather than enhancing the self, while some of them additionally manifested in either an independent mindset/behavior ready for changes or a self-restrictive, preservative mindset/behavior.

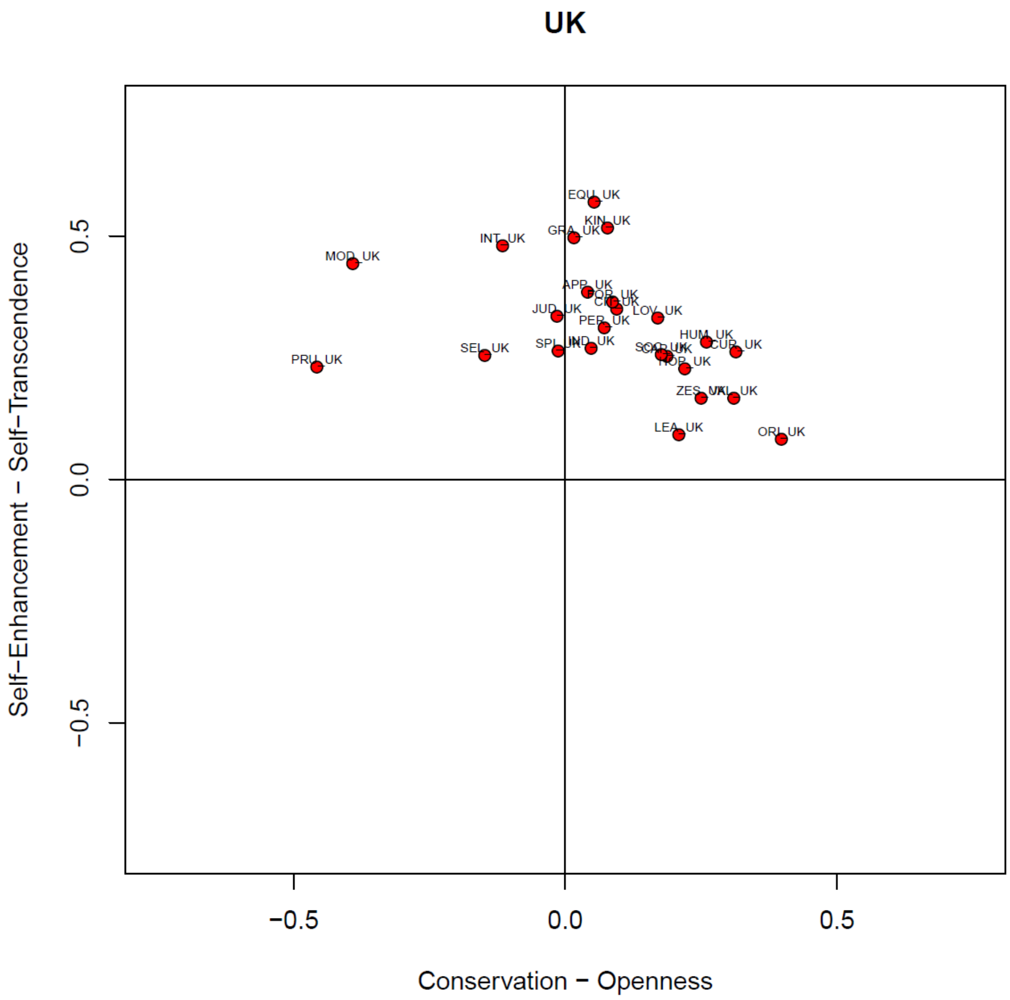


Figure 1.3a. Plot for 24 IPIP-VIA-R scales in the higher-order dimensional space of Schwartz human values in the UK (United Kingdom).

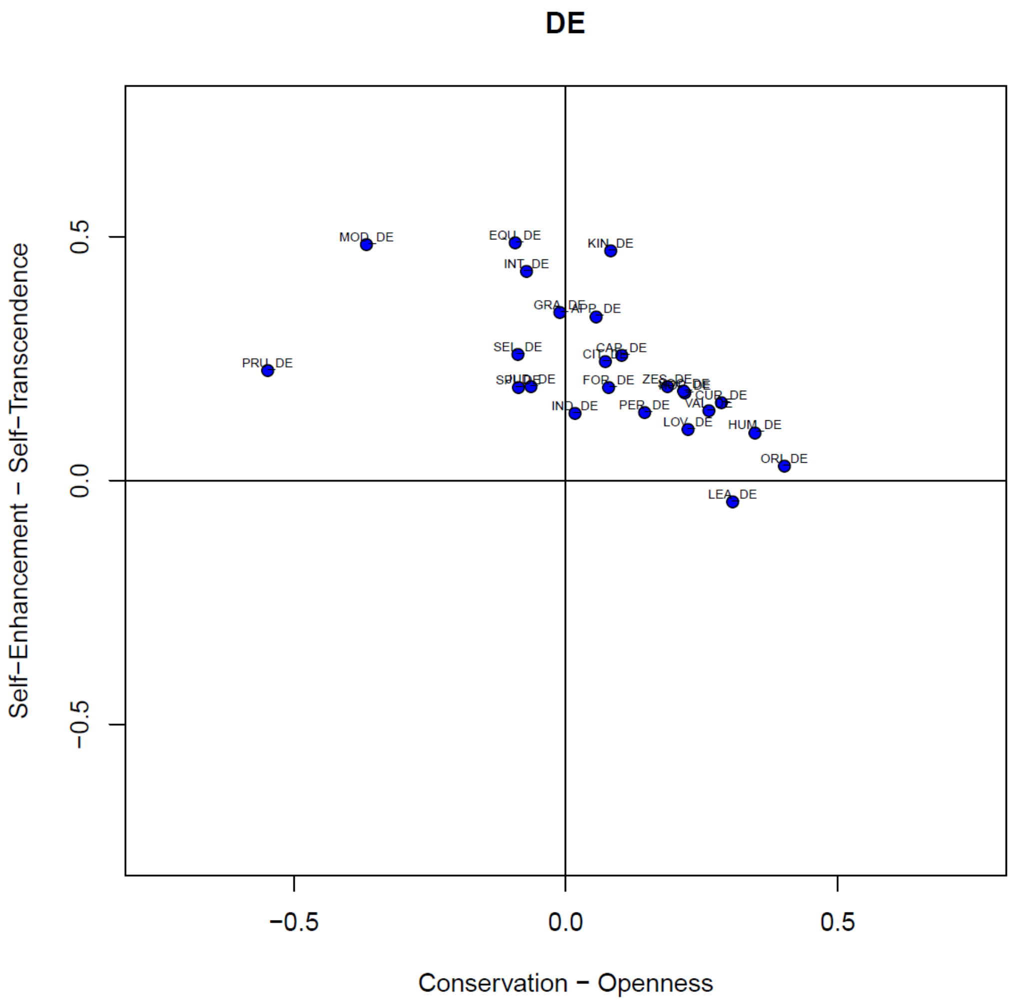


Figure 1.3b. Plot for 24 IPIP-VIA-R scales in the higher-order dimensional space of Schwartz human values in DE (Germany).

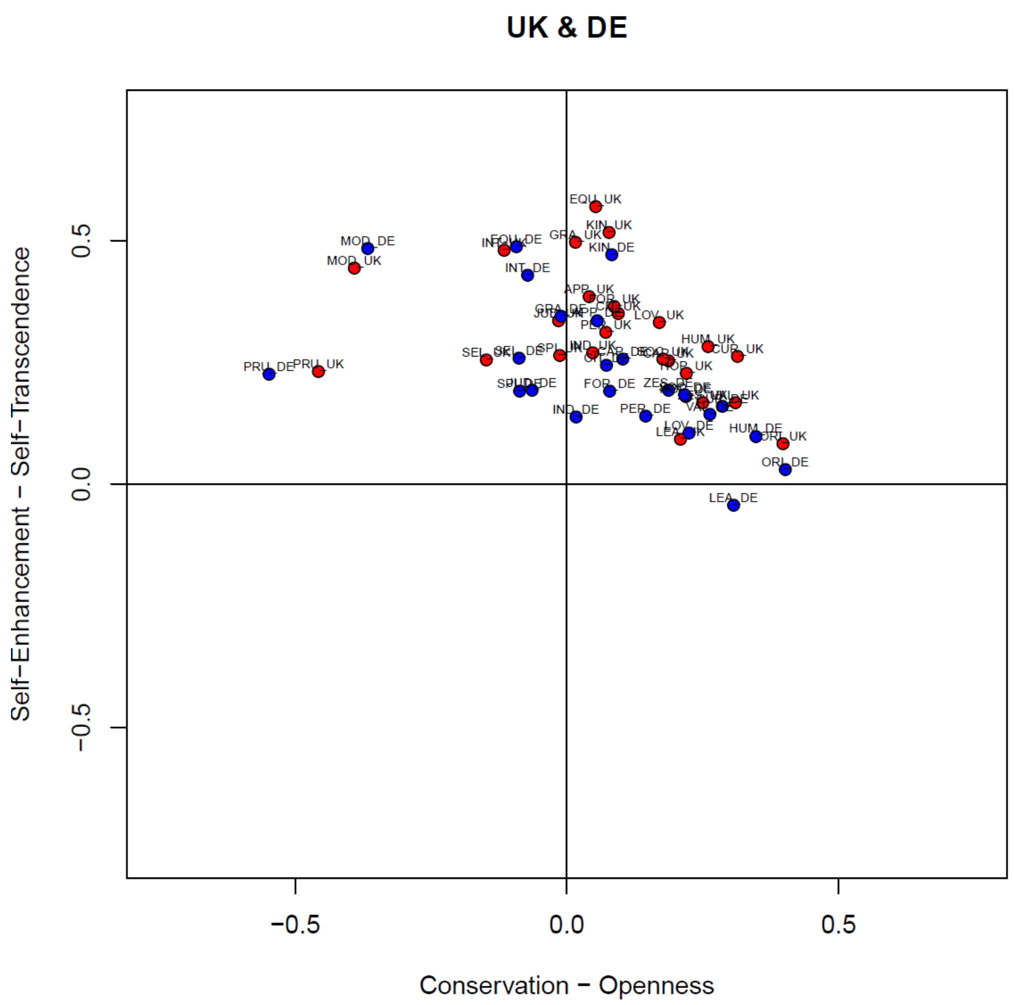


Figure 1.3c. Congruence plot for 24 IPIP-VIA-R scales in the higher-order dimensional space of Schwartz human values across the United Kingdom (UK, red dots) and Germany (DE, blue dots).

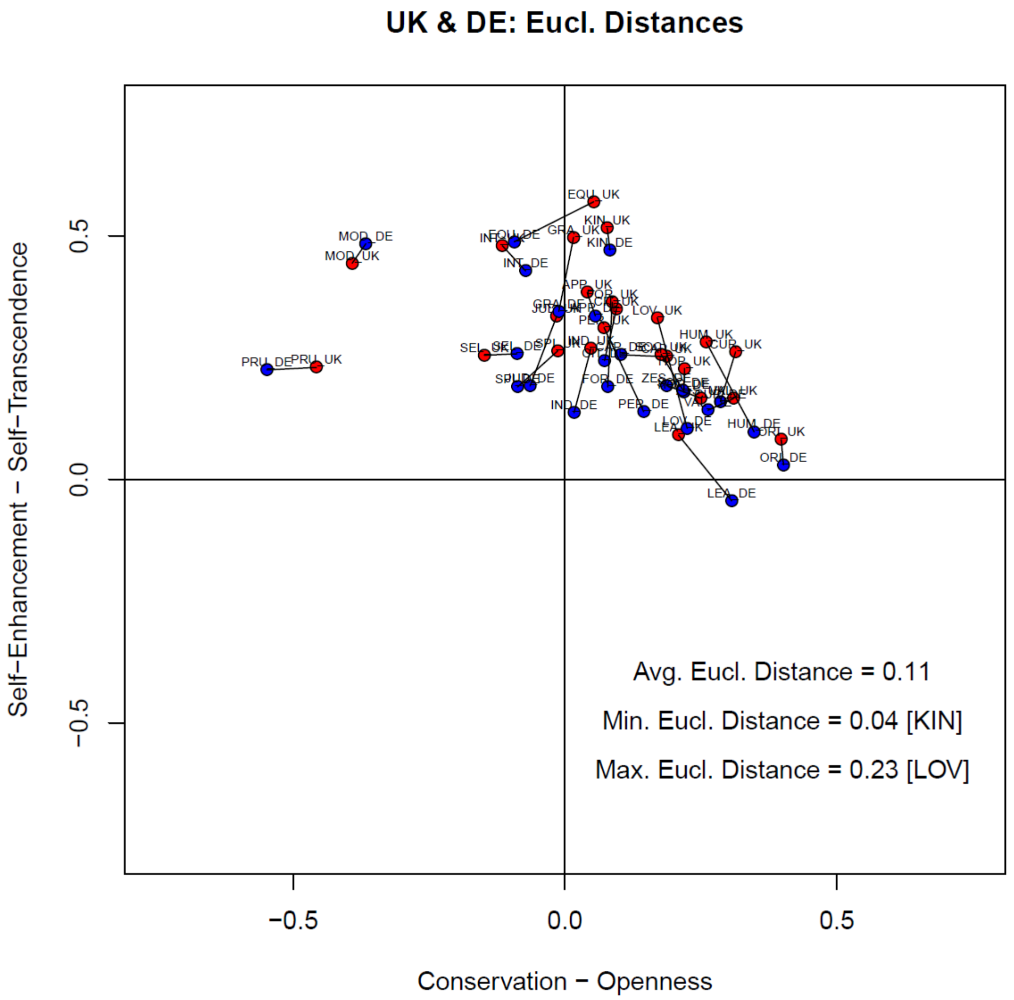


Figure 1.3d. Euclidean distances in the congruence plot for 24 IPIP-VIA-R scales in the higher-order dimensional space of Schwartz human values across the United Kingdom (UK, red dots) and Germany (DE, blue dots). Cross-country differences are indicated by solid lines.

Discussion

Taken together, the results in Tables 1.11–1.13 and Figure 1.3 reveal four important patterns. First, despite noticeable overlap between some character strength scales (for which social desirability may partly account), their pattern of correlations with personality and values provided first evidence that, within the VIA classification the 24 character strengths as measured with IPIP-VIA-R are (largely) non-redundant.

Second, nearly all the strengths shifted markedly towards the Self-Transcendence pole, in line with Peterson and Seligman's (2004) conception of strengths as socially valued traits. From this point of view, being of "good character" means overcoming one's selfish orientations in life. Some strengths closest to the Self-Enhancement pole (or, more precisely, close to the origin of this higher-order value dimension) were also the *least moral* strengths, but at the same time highly valuable for society's and one's own goals: Leadership, Originality, and Industry/Perseverance.

Third, regarding the question of VIA-Big Five redundancy, the answer needs to be differentiated. Each strength was meaningfully related to at least one Big Five personality domain, in most cases to the ones we predicted, and in most cases to all the Big Five domains (although to varying degrees). While this pattern overall supports the IPIP-VIA-R's breadth and nomological network, the multiple associations show that character strengths are interstitial constructs that occupy the intersections of several personality traits. In most cases, we were able to single out one primary personality facet to which each strength was most strongly related, too. At times, the content of these personality facets corresponded closely to the content of the strength scales, sometimes even with the labels being close themselves (compare VIA's Originality to Creative Imagination or VIA's Industry/Perseverance to Conscientiousness's Productiveness facet). Other strengths (e.g., Prudence and Spirituality/Religiousness) showed only a small overlap with Big Five personality domains or facets. Apart from these farthest ends of the spectrum of overlap, most strengths showed (absolute) correlations between *r* = .40 and .60 with the Big Five framework (sans disattenuation).

McGrath and colleagues (2020) reasoned that uncorrected correlations above .60 might serve as evidence for likely scale redundancy, whereas lower correlations of .50 would rather reflect closely related but distinct constructs. If we set the findings for Germany as the reference and apply these criteria to our study, seven character strength scales showed strong overlap with Big Five domains questioning their non-redundancy: Hope, Zest, Industry/Perseverance, Originality, Leadership, Equity, and Kindness. At the proper level of abstraction, strong overlap with facets emerged virtually for the same strengths, with Integrity replacing Equity in the list. All in all, most character strengths are *not* redundant to Big Five domains or facets. They complement the set of personality-related concepts that have been omitted in the Big Five tradition, which focused on person descriptors not subject to moral evaluation. With confidence, we virtually draw the same conclusion as others, but base our judgment on a set of content-valid and cross-culturally invariant IPIP-VIA-R short scales: 17 of 24 strength scales "were not fully accounted for by personality facets" (McGrath et al., 2020, p. 132).

Finally, the correlation coefficients were remarkably similar across the countries, supporting the robustness of the nomological net and the quality of the translation and adaptation achieved. For instance, the average Euclidean distance between UK and German coordinates in the Schwartz value space (see Figure 1.3d) amounted to 0.11. That is, across both axes, the two countries' correlation coefficients deviated hardly at all (or only by 0.11 points on average). Given that we would expect similar relationships between character strengths, personality traits, and values across the culturally similar countries under consideration and together with the encouraging results about scale reliability (see RQ4) and (partial) scalar measurement invariance (RQ5), the overall similar pattern of scale correlations suggests that the IPIP-VIA-R scales work similarly in Germany and the UK.

In sum, refining the IPIP-VIA measures yielded short scales with clear, value-laden, and theory-conforming validity patterns. IPIP-VIA-R poses a valid window into traits that emerged from, and conform to, a cross-cultural perspective. Character strengths complement our view of the personality variable space and look at it from a different angle. Researchers may differ in their preferences for using a theoretically derived, cross-cultural measurement approach (character strengths) or an inductive approach that reflects lexical parsimony (Big Five framework). The next section addresses the utility of these scales by inspecting criterion validity.

Research Question 7:

Is the Criterion Validity of IPIP-VIA-R Scales Sufficient?

Having established the position of the IPIP-VIA-R in a nomological net with personality traits and basic human values, we next tested the scales' criterion validity. Criterion validity is one of the cornerstones of a scale's practical utility. The idea that character strengths promote "the good life" with an optimal balance between pursuing individual life goals and meeting social demands is foundational to the VIA model (Peterson & Seligman, 2004; Seligman, 2002, 2011). Therefore, life satisfaction is a key outcome of which the character strengths should be predictive. To provide a more complete picture of the predictive utility of the IPIP-VIA-R scales for indicators of "the good life", we also assessed respondents' self-reported health. Keep in mind, though, that according to their theoretical conception each character strength is essentially "[…] morally valued in its own right, *even in the absence of obvious beneficial outcomes*" (Peterson & Seligman, 2004, p. 19, emphasis added).

Previous research suggests that positive correlations exist between all character strengths and life satisfaction (Ruch et al., 2007), with the strongest correlations emerging for the transcendence strengths Hope, Gratitude, and Spirituality/Religiousness, and furthermore with the strengths Capacity for Love, Curiosity, and Zest (Park et al., 2004; Proyer et al., 2011; Weber et al., 2013). As alluded to above, we expected positive, albeit somewhat weaker, associations for some strength scales with health: Proyer et al. (2013) found significant positive correlations between a single-item health self-rating and the strengths Hope, Humor, Industry/Perseverance, Self-Regulation, and Zest.

Furthermore, we provide a first exploratory overview of the *incremental* validity above and beyond the Big Five domains of the 24 IPIP-VIA-R scales for life satisfaction and health. We report incremental validity beyond Big Five domains, because the IPIP-VIA-R scales might often be applied next to short scales assessing the Big Five domains rather than the facets (e.g., in large-scale assessment). At the same time, strengths and facets reside on a more comparable level of abstraction or aggregation in the hierarchy of constructs (i.e., the IPIP-VIA-R strengths are conceptually roughly as broad as the Big Five facets) and the number of predictors multiplies in a facet-based analysis. Testing incremental validity of the character strengths beyond the Big Five facets thus represents a stricter approach. While we focus on domain-based incremental validity, we also run facet-based analyses as a cross-validation for domain-based findings (see SOM_IPIP-VIA-R_RQ7_Incremental-Validity.pdf).

Gender. We computed the IPIP-VIA-R correlations with gender, too. Gender may be differentially associated with character strengths. However, the character strengths framework does not provide firm clues for specific hypotheses. Associations between some strengths and gender seem intuitive (e.g., one might expect females to score higher on Kindness; Ruch et al., 2007). Yet, any emerging correlation coefficients need to be interpreted with caution as their origin would be unclear. Large gender difference would alert us to reinspect the criterion validity of strengths only after controlling for gender as a covariate in regression models.

Method

We conducted all analyses in both the UK and Germany using Samples 4 and 5. Previous research had regressed life satisfaction on all strengths in a single step in regression analysis and found that more than 40% of variance overlapped with the criterion (Noftle et al., 2011). For a more nuanced view, we instead investigated each strength's association in separate models. To this end, we measured the general and current satisfaction with life (L-1 scale; Nießen et al., 2020): "All things considered, how satisfied are you with your life these days?" with 11 graded response options ranging from *not at all satisfied* (1) to *completely satisfied* (11). The L-1 scale converges highly with the late Diener's SWLS scale (Diener et al., 1985). We measured health with an *ad hoc* single-item reading "How is your health in general? Would you say it is …" on a 5-point response scale from *very good* (1) to *very bad* (5).

We based gender associations and criterion validity for the two outcomes on bivariate (Pearson) correlations. For analyzing *incremental* validity beyond the Big Five, we conducted multiple-regression analyses like those by Noftle and colleagues (2011). In a baseline model, we first regressed the criterion on all Big Five scores (i.e., five domains or 15 facets). We then added one character strength to the model and judged its incremental validity by its associated ΔR^2 and standardized regression coefficient. Similar to Noftle and colleagues (2011), we also provide the criterion variance explained by each strength (r^2) when regressing the outcome variables exclusively on that strength. Using personality-specific, meta-analytically derived guidelines (Gignac & Szodorai, 2016), we interpreted r = .10, .20. and .30 as small, medium, and large effect sizes, respectively. Accordingly, we interpreted ΔR^2 = 1% as small, ΔR^2 = 4% as medium, and ΔR^2 = 9% as large R^2-change.

Results

We begin by inspecting gender effects. Then we present results on criterion validity of the 24 IPIP-VIA-R scales in Table 1.14 and incremental validity in Table 1.15, respectively (for further details, see also Table D1 in SOM_IPIP-VIA-R_RQ7_Incremental-Validity.pdf).

Gender. For both countries, we mostly found negligible gender differences, though significant positive correlations of small size to medium size emerged with

Spirituality/Religiousness, Kindness, Gratitude, Appreciation of Beauty and Excellence, Integrity, Equity, and Capacity for Love (in descending order). In Germany, Prudence too correlated with gender. Positive correlations indicate that a strength was more pronounced in women. This pattern of correlations is plausible and indicates that, for a few scales, direct comparisons across gender groups may profit from using gender-specific norms. Yet, strongly biased criterion correlations are unlikely. The effect sizes we found are in line with the maximum gender differences observed for the BFI-2 domains (e.g., r_{max} = .17/.26 for Agreeableness in Internet/Student samples; Soto & John, 2017a) and below the meta-analytically derived maximum effect size for Big Five facets $\bar{r}_{max}$ = .24 for Assertiveness (Feingold, 1994).

Table 1.14

IPIP-VIA-R: Gender Associations and Criterion Validity of Character Strengths (Samples 4 and 5)

	Gender		Life Sat.		Health	
Scale	UK	DE	UK	DE	UK	DE
APP	.15	.16	.10	**.29**	.05	**.22**
CAP	.11	.13	**.38**	**.42**	.16	**.22**
CIT	.16	.01	**.24**	**.31**	.08	.19
CUR	.05	.07[a]	**.32**	**.37**[a]	.14	**.31**[a]
EQU	.15	.12	.15	**.22**	.02	.07
FOR	.04	−.07	**.29**	**.27**	.09	.17
GRA	**.20**	.13	**.24**	**.40**	.03	**.24**
HOP	−.01	−.07	**.45**	**.46**	.14	**.24**
HUM	.08	−.04	**.29**	**.34**	.11	**.22**
IND	.05	.02	**.35**	**.40**	.16	.18
INT	.15	.13	.13	.13	.02	−.02
JUD	.00	.04	.15	.16	−.02	.13
KIN	.19	.14	.19	**.25**	.03	.13
LEA	−.01	−.03	**.26**	**.32**	.07	**.25**
LOV	.12	.02	.15	**.30**	.05	**.27**
MOD	.10	.05	−.12	−.05	−.10	−.18
ORI	−.07	.03	.19	**.23**	.06	**.21**
PER	.02	.06	**.25**	**.37**	.09	**.27**
PRU	.05	.18	−.03	−.05	−.08	−.13
SEL	−.12	.02	.19	.14	.13	.13
SOC	.13	.08	**.29**	**.27**	.16	.19
SPI	.16	**.20**	**.35**	**.22**	.03	.12
VAL	−.09	−.04	.16	.15	.00	.14
ZES	.03	−.02[a]	**.54**	**.59**[a]	**.33**	**.36**[a]
Mean	**.07**	**.06**	**.23**	**.27**	**.07**	**.17**
N	474	476	474	476	474	476

Notes. [a] *N* = 475, UK = United Kingdom, DE = Germany; Scale = IPIP-VIA-R strength, Life Sat. = Life Satisfaction; for all coefficients, $p < .05$ if $|r| > .09$; $p < .01$ if $|r| > .12$, $p < .001$ if $|r| > .16$; bold font type: $r \geq .20$ (at least medium effect size).

Table 1.15

IPIP-VIA-R: Incremental Criterion Validity of Character Strengths beyond Big Five Domains (Samples 4 and 5)

	Life Satisfaction						Health					
	UK (N = 474)			DE (N = 476[a])			UK (N = 474)			DE (N = 476[a])		
M0: BFI	R^2_{adj} = .31 (p < .001)			R^2_{adj} = .32 (p < .001)			R^2_{adj} = .16 (p < .001)			R^2_{adj} = .14 (p < .001)		
M1: +Scale	r^2	ΔR^2_{adj}	β	r^2	ΔR^2_{adj}	β	r^2	ΔR^2_{adj}	β	r^2	ΔR^2_{adj}	β
APP	.01	.00	−.03	.08	**.01**	**.15**	.00	.00	.03	.05	**.01**	**.14**
CAP	.15	**.02**	**.20**	.17	**.04**	**.24**	.03	.00	.05	.05	**.01**	**.14**
CIT	.06	.00	−.04	.10	**.01**	**.10**	.01	.00	−.10	.04	**.01**	**.10**
CUR	.10	.00	.07	.14	**.01**	**.14**	.02	.00	−.03	.10	**.02**	**.18**
EQU	.02	.00	−.06	.05	.00	.06	.00	.00	−.01	.01	.00	.02
FOR	.09	.00	.07	.07	.00	.06	.01	.00	.00	.03	.00	.08
GRA	.06	.00	.05	.16	**.06**	**.29**	.00	.00	−.04	.06	**.03**	**.21**
HOP	.21	**.01**	**.16**	.21	.00	.11	.02	**.01**	**−.13**	.06	.00	−.04
HUM	.08	.00	.03	.12	.00	.07	.01	.00	−.05	.05	.00	.02
IND	.12	**.01**	**.15**	.16	**.01**	**.13**	.02	.00	.02	.03	.00	.03
INT	.02	**.01**	**−.12**	.02	.00	−.08	.00	.00	−.06	.00	**.01**	**−.12**
JUD	.02	.00	−.06	.02	.00	.01	.00	**.01**	**−.15**	.02	.00	.05
KIN	.04	.00	−.05	.06	.00	.01	.00	.00	−.06	.02	.00	.02
LEA	.07	.00	−.10	.10	.00	.01	.01	**.04**	**−.29**	.06	.00	.04
LOV	.02	.00	.04	.09	**.01**	**.11**	.00	.00	.00	.07	**.01**	**.14**
MOD	.01	**.02**	**−.17**	.00	.00	−.08	.01	.00	−.02	.03	**.01**	**−.14**
ORI	.04	.00	−.03	.05	.00	.00	.00	**.01**	**−.15**	.05	.00	.00
PER	.06	.00	−.05	.13	.00	.07	.01	.00	−.10	.07	.00	.11
PRU	.00	.00	−.03	.00	.00	.03	.01	.00	−.02	.02	.00	−.01
SEL	.04	**.00**	**−.09**	.02	**.01**	**−.11**	.02	.00	.00	.02	.00	.05
SOC	.08	.00	−.03	.07	.00	.04	.03	.00	−.05	.04	.00	.03
SPI	.12	**.03**	**.19**	.05	**.01**	**.12**	.00	.00	−.06	.02	.00	.06
VAL	.03	.00	−.02	.02	.00	−.07	.00	**.02**	**−.17**	.02	.00	−.01
ZES	.29	**.05**	**.32**	.35	**.07**	**.40**	.11	**.02**	**.20**	.13	**.03**	**.29**

Notes. [a] N = 475 for CUR and ZES; UK = United Kingdom, DE = Germany, Scale = IPIP-VIA-R strength; M0: Baseline model for each criterion and each country only comprising the scale scores of the five Big Five domains as predictors, R^2_{adj} = multiple R-squared (adjusted) of each baseline model; M1: Model additionally including scale score of IPIP-VIA-R strength as predictor, r^2 = squared bivariate correlation of IPIP-VIA-R strength with criterion, ΔR^2_{adj} = difference in multiple-R squared (adjusted) between M0 and M1, β = standardized regression coefficient of IPIP-VIA-R strength; bold font type: p < .05.

Criterion Validity: Life Satisfaction. In line with theory, in comparison with gender and health, life satisfaction showed the highest average correlation with character strengths ($\bar{r}_{UK}$ = .23, $\bar{r}_{DE}$ = .27). In line with previous findings, large correlations were found with Hope, Gratitude, Spirituality/Religiousness, Capacity for Love, Curiosity, and Zest either in both the UK and Germany or in at least one of the countries. In addition, across both countries, life satisfaction had large correlations with Industry/Perseverance. In Germany, additional five correlations were of large magnitude. All other correlations were positive and ranged between small and medium/large size. The only exceptions were the correlations of life satisfaction with Modesty (non-significant in Germany, and negative in the UK) and Prudence (statistically non-significant in both countries). From a theoretical point of view suggesting that *all* character strengths should somehow contribute to a happy life, the findings for the latter scales are somewhat unexpected, yet not completely implausible. Modesty may be a “mixed blessing” in that it may prevent return of investments in social or economic respects. Also, people scoring low on Modesty might want something special out of life and might consider themselves as satisfied as highly modest people. People scoring high on Prudence might hesitate to evaluate their lives as satisfying, either because they focus on risks, undermining their satisfaction, or because life has been rife with disappointments, making them cautious. Also, cultures may differ on how much importance is given to personal happiness as a criterion for a “good life.” Modest and prudent people may be orienting their lives more toward something other than self-gratification.

In both countries, Capacity for Love, Industry/Perseverance, Spirituality/Religiousness, and Zest showed positive incremental validity for life satisfaction above and beyond the Big Five domains. Also, Curiosity, Gratitude, and Hope, which appeared as strongest correlates of life satisfaction in previous and present studies, showed incremental validity beyond the Big Five in one of the countries. The increment for Self-Regulation had a negative beta-weight in both countries, indicating that the positive bivariate correlations were generated by parts that overlapped with the Big Five, and the unique part predicted life satisfaction negatively, potentially indicating affective-behavioral over-controlling. Other incremental effects varied across countries (see Table 1.15). Among the significant incremental effects, the R^2-change ranged between 0.4–5% in the UK and 1–7% in Germany, that is, between small and medium to large size.

In the facet-based cross-validation—based on powerful three times the number of predictors in the baseline model—the 10 significant incremental effects for predicting life satisfaction in the domain-based model in Germany replicated. In the UK, for life-satisfaction six out of eight increments beyond domain scores replicated with facets. Effect sizes were smaller on average, but still substantial R^2-changes of 0.4–5% for life satisfaction (for details, see SOM_IPIP-VIA-R_RQ7_Incremental-Validity.pdf).

Criterion Validity: Health. The average correlation between the 24 IPIP-VIA-R scales and health varied across the two countries: Whereas in Germany most correlations ranged between .10–.30, in the UK more than half of the correlations were non-significant. In line with Proyer and colleagues (2013), we identified Hope, Humor, Industry/Perseverance, Self-Regulation, and Zest as significant correlates of health across both countries: Hope, Humor, Industry/Perseverance, and Self-Regulation were among the scales with small to medium-sized (significant) correlations in UK and produced mostly mid-size correlations in Germany. In both countries, Zest even showed a large correlation. Noteworthy, Curiosity was also a large positive correlate in Germany. Prudence showed a significant negative correlation with health in Germany, as did Modesty in both countries. This finding is compatible with the view that Prudence and Modesty may not be universally beneficial, although the cross-sectional data cannot rule out that bad health may conduce to the development of both "strengths", Prudence and Modesty, or that these strengths are sometimes formed as a result of negative life experiences that restrict life satisfaction and health, which in turn may cloud the otherwise positive contribution of these strengths to life satisfaction and health.

Given that many character strengths were uncorrelated with health, incremental validity of IPIP-VIA-R scales beyond the Big Five was limited. Also note that even all the Big Five taken together hardly explained large amounts of criterion variance (≤ 16%). On the positive side, Zest possessed incremental validity in both countries, despite its strong overlap with Neuroticism (and the BFI-2-S's Depression facet in particular). Furthermore, in Germany, about one quarter of the IPIP-VIA-R scales showed both significant positive correlations with health and positive incremental validity. Modesty not only correlated negatively in Germany, but also had a negative beta-weight. Though this finding was not corroborated cross-culturally, it leaves open the possibility that unique parts in Modesty that do not overlap with the Big Five are rather detrimental to health, undermining a strict interpretation of this trait as character *strength*. Unfortunately, most incremental effects found in UK cannot be interpreted meaningfully (content-wise), because the bivariate correlations were already non-significant (and negative beta-weights are likely due to statistical suppression). With R^2-change ranging between 1–3%, the incremental effects in Germany were small (or medium at best), as was the incremental effect of Zest in the UK (2%).

In the facet-based analysis in Germany, seven out of nine incremental effects mirrored the findings from the domain-based analyses (with R^2-changes of 1%). In contrast, the only well-interpretable effect predicting health beyond domain scores in the UK, Zest, did not replicate with facets.

Discussion

Our findings corroborated the criterion validity of the VIA-IPIP-R scales for life satisfaction and health. On average, we replicated across countries previously reported large and medium positive associations of strengths with life satisfaction and health, respectively. More (and descriptively stronger) positive correlations across the criteria emerged in Germany than in the UK. Rather than speculating about stronger predictive validity in a German context, we attribute the differences to better data sample quality (for which we have independent indications such as unit-nonresponse) and the favorable German translation/adaptation. At any rate, the robust and occasionally large correlations of the character strengths with the criteria demonstrate that the IPIP-VIA-R scales not only have good structural validity (see RQ2–4) but also predictive utility. This is especially true for life satisfaction—the outcome with which strengths should be associated if they indeed promote "the good life", as stipulated by the VIA framework (Peterson & Seligman, 2004), even if character strengths are usually morally valued in their own right (Stahlmann & Ruch, 2020).

We also found some evidence that IPIP-VIA-R has incremental validity above and beyond the Big Five domains. Because there is a clear conceptual overlap between the VIA framework, Big Five domains and especially some facets (see RQ6 and Table 1.13), our analyses constitute a very conservative test of the character strengths' incremental predictive validity. Nonetheless, several strengths showed incremental validity for the considered criteria beyond the Big Five domains (Table 1.15). Life satisfaction emerged as the criterion associated most strongly with character strengths beyond the Big Five, whereas health was harder to predict incrementally and overall.

Correspondingly, some of the resulting effect sizes were medium to large for the criterion of life satisfaction, though smaller for health. We interpret the incremental value of character strengths (R^2-change) against the background of the baseline models (R^2) in which the Big Five domains were the only predictor variables: About 30% and 15% of criterion variance were explained by the Big Five domains in life satisfaction and health, respectively. Adding character strengths explained 4–7% more of the variance at most, which we consider substantial, given that a single character strength then has about the same influence as has the average Big Five domain. In Germany, most incremental effects replicated when using multiple Big Five facets instead of domain scores. Unfortunately, the same cannot be said to apply to the UK, which might be attributed to the alleged differences in sample quality and improved item wordings in the German adaptation of IPIP-VIA-R items. In sum, our results demonstrate the utility of IPIP-VIA-R for predicting important life outcomes—sometimes over and above the Big Five framework, regardless whether using domain or facet scores.

General Discussion

About two decades ago, Peterson and Seligman (2004) presented a seminal classification of character strengths that has led to a renaissance of "human character" as empirical research topic. The concept of human character has interested philosophers for millennia but had long been abandoned in personality psychology in its fundamental search for personality characteristics that arose from social evaluation. In contrast to the lexical Big Five tradition in which evaluative (including moral) adjectives were deliberately, though only partially, removed from personality item pools, Peterson and Seligman's (2004) VIA framework is based on a comprehensive theoretical analysis of personal qualities that they deemed (morally) "valued in their own right" across cultures and that they thought promoted individual and collective well-being, or in short: the "good life." Owing to its cross-cultural claim and exclusive focus on evaluative traits, the VIA framework can provide a valuable complement to the Big Five or HEXACO models for individual differences research. It is therefore not surprising that researchers from different disciplines have embraced the VIA framework and produced a rapidly growing body of evidence on the precursors and outcomes of character strengths.

However, as our review showed, limitations in the VIA measurement instruments most widely used for many years threaten the validity of much of the extant research on character strengths. While proprietary VIA survey instruments were recently revised (e.g., McGrath et al., 2022; McGrath & Wallace, 2021), the limitations persist in the public domain IPIP-VIA inventory. In the present study, we therefore reported on a major research project through which we aimed to advance the assessment of human character in the public domain by developing short, balanced-keyed, content-valid, and cross-culturally applicable scales for the 24 VIA character strengths based on the open-access IPIP.

Based on the findings presented, we can now answer the seven research questions that guided our project. (1) As we suspected, and as previous research has pointed out for VIA-IS items (Ng et al., 2017), the original 252 IPIP-VIA items do not allow for unidimensional assessment of each character strength. Moreover, as the often small loadings showed, many items have little in common with the core of their target strength, meaning that they lack validity. (2) Fortunately, however, the IPIP does comprise enough content-valid, cross-culturally applicable, and well-worded items so that balanced-keyed short scales comprising four items per strength can be constructed. The resulting scales largely preserve the substantive breadth of character strength definitions but are nonetheless essentially unidimensional. Remaining deviations from strict unidimensionality are likely due to ARS, which constitutes an additional, content-unrelated source of variance.

(3) Single-factor CFA measurement models for each of the 24 character strength scales that control for acquiescent responding largely show good fit and mostly satisfactory factor loadings. Three scales—Hope, Leadership, and Spirituality/Religiousness—may require refinement in the long run to minimize correlated residuals that deviate from essential unidimensionality. (4) As to whether the IPIP-VIA-R short scales measure reliably, the answer is a qualified yes. For applied scale use, we recommend omega from adequately specified measurement models with ARS-control, alternatively test-retest reliability for manifest scale scores. While for individual diagnostics the scales might have to be amended with more (content-valid) items, the reliability figures were sufficient for the purpose of social surveys. Random measurement error does not bias the (unstandardized) regression coefficients when a scale is used as an outcome in such a survey (Lechner et al., 2021). When character strengths are used as *predictors* of life outcomes, measurement error can be controlled by using latent measurement models (such as the one presented in RQ3).

(5) Across two languages, all 24 scales performed at the metric invariance level (except for Gratitude where a sole item was found to be non-invariant). This pattern speaks to

the same use of measurement units and psychological meaning of strength items, which is a requirement for cross-group comparisons of factor variances, covariances, and validity correlations (or regression approaches). Moreover, all scales even reached full or partial scalar invariance. (Partial) scalar invariance is a requirement for comparing latent means of strength scales across countries. Scalar invariance is often a hard-to-reach goal in cross-cultural research. In view of this, the high degree of measurement invariance the IPIP-VIA-R scales reached is very encouraging and is testament to the success of the rigorous item selection and translation/adaptation process of our project, in which cross-cultural applicability was one of the guideposts from the outset.

(6) Analyses of nomological nets showed that the 24 character strength scales had theoretically plausible associations with both personality domains and facets and basic human values and supported both the convergent and discriminant validity of the character strengths. Analyses suggested that character strengths overlap with both personality traits and values, which makes sense because character strengths are morally valued personal qualities. Almost all character strengths reflected Self-Transcendence values (Schwartz, 1994, 2003a), supporting the claim that they are not primarily in the service of individual self-enhancement but communal in nature. Few character strengths were so strongly related to (sometimes even similarly named) personality facets that the constructs must be deemed identical. This is not surprising, given that VIA attempted to comprehensively compile positively valued character traits and that character strengths partly use the same person-descriptive adjectives as the lexical tradition. At the same time, there were several strengths (e.g., Spirituality/Religiousness, Prudence) that had no conceptual counterpart in Big Five personality domains or their underlying facets at all, and most strengths had substantial unique portions of variance that they did not share with personality traits and basic human values. Thus, character strengths occupy their own position in the trait space spanned by personality traits and values.

(7) We found initial evidence in support of criterion validity. Specifically, the IPIP-VIA-R character strength scales predicted life satisfaction and health. Even evidence for incremental validity beyond the Big Five, albeit small, emerged. The verdict is a bit more optimistic for the German than the English scales due to the quality of wordings achieved in the translation/adaption process. Objective indicators are needed to corroborate these findings.

Our findings show that it is possible to measure character strengths as conceived in the VIA framework with IPIP. We are confident that IPIP-VIA-R currently represents the best solution for each character strength that can be obtained within the IPIP alone. The IPIP-VIA-R is a major advancement over previous VIA instruments that were predominantly used for a long time in that it is explicitly designed with content validity and cross-cultural applicability in mind, in line with the cross-cultural theoretical underpinnings of the VIA framework. Our findings suggest that the joint rational-empirical approach to deriving brief balanced IPIP-VIA-R scales was successful and yielded efficient, reliable, and valid strength measures.

Limitations and Future Directions for Character Assessment

Despite the improvements over other character inventories based on VIA, and especially the original IPIP-VIA, IPIP-VIA-R is not perfect. In this regard, our self-imposed constraint to use exclusively items from IPIP has some drawbacks. Although it was our preset goal to measure each character strengths with four items (similar to, for example, the BFI-2 facet scales; Soto & John, 2017a), it was sometimes challenging to select even four items from IPIP-VIA that met all our criteria of content validity, item wording, and cross-cultural applicability (see RQ2). Occasionally, we had to resort to IPIP items from other sections than IPIP-VIA to meet all demands. Although four items are sufficient for research purposes—which is what IPIP-VIA-R is designed for—the present IPIP-VIA-R inventory is not sufficiently reliable for individual diagnostics, and IPIP does not provide enough valid items to construct longer IPIP-VIA scales that would be suitable for individual character assessment. Moreover, the wording of some IPIP items is less than optimal. Items were invented, often by local project volunteers, mostly from the United States. Also, the IPIP limits the options for finding suitable items representing unique aspects of VIA character strengths, because content substantially related to the Big Five core is heavily represented, despite later attempts to shape the IPIP item set somewhat into the direction of content less related to the Big Five (Saucier & Goldberg, 1998). Yet, by rationally evaluating item content according to strength relevance and wording clarity, we promoted focus and fidelity of each strength scale. While minimizing the influence of ARS by balanced item keying, we also balanced bandwidth and fidelity as much as the IPIP framework allowed without inflating reliability by redundant or multifactorial item sets, but by focusing on the common core of each strength.

Note that we only investigated two Western countries and languages. Developing IPIP-VIA-R in parallel in two languages was a major advantage and proved fruitful because we considered cross-cultural applicability and translatability already during item selection (see RQ2). Future translations of IPIP-VIA-R are needed for surveys that transcend cultural settings beyond the Indo-Germanic languages we investigated. We believe that the IPIP-VIA-R provides a better starting point for future translations than any prior VIA instrument and would recommend the German over the English version as a source because, as noted in RQ2, we slightly improved the wording of the German translation during the adaption process.

We hope that the IPIP-VIA-R short scales provide an apt measurement basis for future studies, enabling economic assessment in large-scale surveys and longitudinal follow-up. We reiterate desiderata from a decade ago by Noftle and colleagues (2011). Future research needs to address open questions about the validity of the VIA framework and, to this end, pursue long-term real-world outcomes. Longitudinal studies that examine character development over extended periods of time are required. Research that tackles the validity of VIA character strengths, also with the help of longitudinal designs, has gained new momentum recently (e.g., Gander et al., 2020; Stahlmann & Ruch, 2020; Wagner et al., 2020).

Future investigations might also focus on comparative predictive capacity with other instruments assessing strengths. The IPIP advocates the idea of comparative-validity studies that compare original personality inventories for which IPIP scales have been developed with their IPIP counterpart (https://ipip.ori.org/newResearchOpportunities.htm). With IPIP-VIA-R strength scales being more in line with a cross-culturally comparable unidimensional measurement approach than any version preceding it, the exploration of the hierarchical structure among strength scales has become a worthy endeavor.

Conclusion

Based on a rigorous psychometric evaluation, we recommend the use of IPIP-VIA-R as a basis for future research on character strengths. We believe it works better than existing instruments and it is freely accessible in the public domain—free of charge and without registration.

Endnotes

[1] We based our revision on standard IPIP items, but not newly added (Rasch-modeled) VIA items that were exclusively used in a South African context (variable set W; du Plessis & de Bruin, 2015). Concerns about low sample size, the specific study context (as well as unavailability to an extant US sample), and content validity cast doubts on their utility and the respective Rasch analyses.

[2] We identified careless responders *indirectly*, yet we mention here four interspersed direct attention checks and two diligence items in the questionnaire (as part of an unrelated project on careless responding).

[3] Some scoring keys on the IPIP webpage (https://ipip.ori.org/newVIAKey.htm), which served as the inspirational starting point for our investigation, differ slightly from the scoring keys used on the ESCS data presented on the Harvard Dataverse archive (https://dataverse.harvard.edu/dataverse/ESCS-Data). A description of how those differences arose can be found at https://ipip.ori.org/VIAkeys-explanation.htm.

[4] We do not assume that orthogonality reflects “true” factor structures underlying strengths. We merely used Varimax as a tool to establish the necessary minimum number of factors and their associated variance proportions. These values merely served descriptive statistical purposes (not substantive interpretation). With default settings, SPSS provides no details on factors with initial Eigenvalues < 1 that are not of interest here.

[5] Two Heywood-cases (i.e., an item communality reached or exceeded 100%, or a standardized loading exceed 1.00) occurred. Extracting the number of factors suggested by parallel analysis may be statistically necessary for two item sets (CAP, FOR), but does not yield factors that are sufficiently “overdetermined” by multiple variables (i.e., single items may determine their last factors).

[6] Despite (initially) independent experts, our procedure violates the assumption of independent judgments *within* scales, thus prevents the application of rigid rules and chance-corrections (e.g., Lawshe, 1975).

[7] MRFA minimizes the amount of common variance left unexplained (Ten Berge, 1998; Ten Berge & Kiers, 1991). MRFA is the only method that enables the percentage of ECV to be computed in the common factor model (Lorenzo-Seva, 2013). Other than principal axis factor analysis, which yields non-positive definite reduced matrices, MRFA returns strictly positive semi-definite reduced correlation matrices, so that all Eigenvalues are positive and derived coefficients computable (Lorenzo-Seva, 2013; Shapiro & Ten Berge, 2002). Computing ECV from principal components is not appropriate with few variables and limited communality in short scales, as PCA confounds common with unique variance, so that loadings are inflated (Widaman, 1993). For the same reason, even with correctly specified categorical indicators and polychoric correlation matrices, running component-based MAP tests on short scales with few variables only shows a biasing tendency towards underfactoring (Garrido et al., 2011).

[8] The observed rater agreement for the preferred items prior to negotiation was 79% compared to 62% expected by chance, and it improved to 81% after negotiation (Freelon's 2010, webservice ReCal3).

[9] Using stringent cutoffs rigidly is disavowed because models may get rejected that are, in fact, appropriate (Marsh et al., 2004). Fit heuristics have their merit when tested models reflect the simulation conditions from which cutoffs were derived. However, RMSEA is inflated with a low number of items (e.g., $k = 4$), even when a model is correctly specified (Kenny et al., 2015; Kenny & McCoach, 2003; Shi et al., 2019), and CFI is too low if the null model has some plausibility, that is, when conceptual breadth and item heterogeneity restricts inter-item correlations to some extent (Kenny et al., 2015).

[10] While specifics are beyond the present scope, a cursory inspection of unidimensional models that omitted the ARS index resulted in significant χ^2-tests for *all* Variant A and B scales in the UK (Sample 4) and poor fit, mean CFI = .63, RMSEA = .26, SRMR = .08; similarly, in Germany (Sample 5), mean CFI = .85, RMSEA = .14, SRMR = .04. Dropping ARS for a few scales where model fit might suffice on the surface yields inconsistent measurement models across countries and IPIP-VIA-R scales.

[11] Comparing ω to gHI values (see SOM_IPIP-VIA-R_RQ4_Reliability.pdf) shows that using unit-weighted (ω) instead of ideally weighted composites (gHI) is *usually* not detrimental to the reliability of IPIP-VIA-R short scales. We observed little differences. The maximum discrepancies were .15 for Social Intelligence in Germany, and .21 for Self-Regulation in UK.

[12] From the normative point of view, the strengths should be appreciated as morally good; even if they are not moral strengths in the strict sense, they should form strengths of character that enable success for society's or the greater good. Thus, people of character cannot simply strive towards individualistic goals, and Self-Enhancement cannot be the driver. Character strengths help overcoming egoism, which is relevant for the evolution and longevity of social organisms. For two strengths (Leadership, Industry/Perseverance) the raters spontaneously predicted a positive association with Self-Enhancement. For Conservation versus Openness to Change, we mostly expected the strengths to scatter along this axis (maybe slightly shifted towards Openness; see SOM_IPIP-VIA-R_RQ6_Nomological-Net-Hypotheses.xlxs). We refrained from forming strong hypotheses about IPIP-VIA-R Spirituality/Religiousness scale, as it blends values related to Self-Transcendence/Self-Enhancement and Openness to Change/Conservation. Various constellations of Believing, Bonding, Behaving, or Belonging shape one's religious style or spiritual experiences, and these four aspects are stressed differently across cultural and socio-economic contexts (Gennerich & Huber, 2006; Saroglou, 2011). Whereas Spirituality may

relate to a preference for Openness to Change and Self-Transcendence values, Religiousness may relate to Conservation and Self-Enhancement (e.g., Tradition expressed in rites and beliefs; Peterson & Seligman, 2004; Saroglou et al., 2004; Saucier & Skrzypińska, 2006; Schwartz & Huismans, 1995).

STUDY 2

Abstract

Peterson and Seligman's (2004) values-in-action (VIA) framework maps 24 character strengths onto six more abstract virtues through a theoretical classification. However, compared to other individual difference constructs, there is little consensus about the *factor-analytic* structure of the VIA trait space. Applying Horn's parallel analysis, Goldberg's Bass-ackwards approach, and cross-country congruency analysis, we scrutinize the factor-analytic solutions-hierarchy of the 24 VIA strengths with the aim to identify one or more useful *global* levels of abstraction (akin to the Big Five, HEXACO/Big Six, or personality metatraits). We assessed the 24 character strengths with the psychometrically refined IPIP-VIA-R inventory in two large and heterogeneous samples from Germany and UK (total $N \approx 2{,}000$). Results suggested that three global dimensions suffice to capture the essence of character strengths: Level III recovered more than 50% of the total variation of the 24 character strengths in well-interpretable, global/general, cross-culturally replicable dimensions. We provisionally labeled them *positivity*, *dependability*, and *mastery*. Their superordinate Level-II-dimensions were reminiscent of the "Big Two" personality metatraits *Dynamism* and *Social Self-Regulation*. Our results advance the understanding of the VIA character trait space and may serve as a basis for developing scales to assess these global dimensions.[†]

Keywords: VIA, character strengths, personality structure, metatraits, higher-order factors

[†] Study 2 was published as Partsch, M. V., Bluemke, M., & Lechner, C. M. (2022). Revisiting the hierarchical structure of the 24 VIA character strengths: Three global dimensions may suffice to capture their essence. *European Journal of Personality*, *36*(5), 825–845. https://doi.org/10.1177/08902070211017760. This version is the final author version before copy editing.

Introduction

The question of how to define and describe human character has interested philosophers since antiquity. In modern-day psychological research, Peterson and Seligman's (2004) values-in-action (VIA) classification of character strengths and virtues ranks among the most prominent and well-established approaches to studying "good character". These authors identified 24 theoretically justified and empirically supported character strengths, which they assigned to six more abstract virtues. They developed the Values in Action Inventory of Strengths (VIA-IS) to measure these 24 character strengths (Peterson et al., 2005).

Whereas Peterson and Seligman's (2004) original classification of character strengths to virtues is based on *theoretical* considerations, subsequent studies on individual differences in character have used techniques such as principal component analysis (PCA) or exploratory factor analysis (EFA) to investigate the *empirical* (factor-analytical) structure of the VIA character trait space and to identify higher-level dimensions (i.e., factors or principal components) on which to aggregate the 24 character strengths (e.g., Anjum & Amjad, 2019; Brdar & Kashdan, 2010; Macdonald et al., 2008; McGrath, 2015; Peterson et al., 2008; Ruch et al., 2010; Shryack et al., 2010; Singh & Choubisa, 2010). However, these studies have not led to a consensus regarding the most useful global level(s) of abstraction on which to aggregate the 24 VIA character strengths.[1] Thus—in marked contrast to other major individual difference constructs such as personality traits from the lexical tradition or intelligence—the VIA research tradition still lacks consensus about useful global levels of aggregation (such as "domains" or "metatraits") above the 24 character strengths. The varying nature and quality of the measurement instruments, samples, and methodologies used in previous studies have likely contributed to this unfortunate state of affairs.

In the present study, we revisit the hierarchical structure of the 24 VIA strengths through a rigorous factor-analytic approach. Our aim is to identify one or more global levels in the solutions-hierarchy revealed with the Bass-ackwards approach (Goldberg, 2006) that comprise (1) well-interpretable, (2) global/general, and (3) cross-culturally replicable higher-level dimensions. We measure the 24 character strengths with the IPIP-VIA-R inventory (Bluemke et al., 2021), a selection of 96 items (i.e., four per VIA character strength) from the established International Personality Item Pool (IPIP; Goldberg et al., 2006) based on content validity, unidimensionality, and other psychometric criteria. With the present study, we hope to contribute to a better understanding of the VIA trait space and provide researchers with different levels of abstraction on which to describe individual differences in character for different research purposes.

Theoretical and Empirical Approaches to VIA

The Theoretical Classification of Character Strengths and Virtues

Peterson and Seligman (2004) endeavored to map out the realm of human excellence and describe people of “good character”. Through an extensive theoretical analysis of religious doctrines, philosophical traditions, and cultural artifacts from China, South Asia, and the West (e.g., Confucianism, Buddhism, and Judeo-Christianity), they derived six abstract “core virtues”. They considered the six virtues of *courage*, *justice*, *humanity*, *temperance*, *transcendence*, and *wisdom* to be cultural universals.[‡]

Furthermore, they identified 24 more specific “character strengths” based on an extensive collection of positive traits. Character strengths represent specific instances, realizations, or ways of expressing, these virtues. They represent morally valued character traits that can contribute to a fulfilled life for both the self and others. These character strengths are the theoretically justified and empirically supported (see Peterson & Seligman, 2004) building blocks of the VIA character trait space similar in abstraction to personality

[‡] We describe the six virtues in more detail in the Supplementary Online Material (SOM). We provide this and all other SOMs on the project website at the Open Science Framework (OSF) at https://osf.io/m9aev/.

facets (see Bluemke et al., 2021; McGrath et al., 2020; the tables from Bluemke et al., 2021, in the SOM on OSF).

Peterson and Seligman (2004) assigned each of the 24 character strengths to one of the six virtues. They modeled their theoretical classification on Carl Linnaeus's classification of species, which is based on common attributes. That is, they identified conceptual similarities among the 24 character strengths, based on which they mapped them onto the six pre-defined core virtues (see SOM on OSF for more details on Peterson and Seligman's (2004) classification approach). For example, *valor*, *industriousness*, *integrity*, and *zest* are "emotional strengths that involve the exercise of will to accomplish goals in the face of opposition, external or internal" (p. 29), which map onto the virtue *courage*. They described their classification as a "hierarchy of abstraction".[2]

Peterson and Seligman (2004) maintained that a person possesses a virtue if they show one or two—but usually not all—of the strengths subsumed under this virtue.[3] For example, a person may score low on industriousness and zest but still be considered to possess the virtue of courage if they score high on valor and integrity. Importantly, this implies that character strengths classified under the same virtue as different instances of that virtue are not necessarily correlated with each other.

Empirical Factor-Analytic Approaches to the 24 VIA Character Strengths

Peterson and Seligman's (2004) theoretical classification of character strengths to virtues based on shared attributes must be demarcated from empirical, factor-analytic approaches that are based on observed patterns of correlations among the 24 character strengths. Factor-analytic studies employ Goldberg's (2006) Bass-ackwards approach or related techniques to establish the hierarchical structure of a trait space and identify (mostly global) levels of abstraction on which individual differences can be described. Different from Peterson and Seligman's (2004) theoretical classification, lower-level dimensions in factor-analytic investigations are assigned to higher-level dimensions based on their empirical correlations, expressed through factor loadings. In contrast to the scoring assumptions in Peterson and Seligman's (2004) Linnaean classification, a person scoring high on a factor-analytically derived higher-level dimension in the VIA trait space will tend to score relatively higher on all lower-level dimensions (e.g., the VIA character strengths) associated with that same higher-level dimension.[4]

Although both approaches share the goal of describing human character on different levels of abstraction, the theoretical and empirical (i.e., factor-analytic) approach are ultimately incommensurable, and their conclusions need not coincide. Factor-analytic methods cannot directly test Peterson and Seligman's (2004) theoretical classification of character strengths and virtues because this classification is not based on observed correlations of the character strengths, although there are other methods (e.g., based on expert ratings) to empirically validate this classification (see Ruch et al., 2019; Ruch & Proyer, 2015). Thus, it cannot be expected that the six virtues identified by Peterson and Seligman (2004) will be recovered through factor-analytic means, even though it may of course occur.

While we consider theoretical and factor-analytic approaches to establishing a trait hierarchy as complementary and equally legitimate, our focus in the present paper is exclusively on the latter. In research on individual differences, the factor-analytic approach is a well-established and widely used standard approach to establishing the hierarchy of a trait space. For example, Bass-ackwards analysis and related techniques have been successfully used to unravel the hierarchical structure of intelligence and personality and identify different useful aggregation levels. These aggregation levels reach from global, encompassing only a few (e.g., two to six) highly aggregated traits, to specific, encompassing a large number (e.g., 15–30) of narrow traits. Specifically, the trait hierarchy of intelligence comprises a *G* factor (general mental ability; e.g., Jensen, 1998) that can be broken up into two more specific factors (fluid and crystallized intelligence; Cattell, 1943, 1963), which in turn subsume a range of more specific abilities (e.g., Lang et al., 2016). Similarly, the trait hierarchy of personality ranges from a General Factor of Personality (Rushton et al., 2008; albeit highly controversial and of questionable utility), two metatraits (e.g., Stability and Plasticity as in DeYoung, 2006; or the "Big Two" Dynamism and Social Self-Regulation, as in Saucier et al., 2014), the Big Five domains (McCrae & John, 1992) and the six HEXACO/Big Six domains (Ashton & Lee, 2007; Thalmayer & Saucier, 2014), followed by more fine-grained levels such as aspects (DeYoung et al., 2007), facets (Ashton & Lee, 2007; Costa & McCrae, 1992; Soto & John, 2017a), and nuances (Mõttus et al., 2017).

Establishing the trait hierarchy of individual difference constructs and identifying useful levels along the hierarchy's continuum of abstractness/generality is important for several reasons. First, it helps to better understand the nature and makeup of the construct. For example, how many global higher-level dimensions span the VIA trait space populated by the 24 VIA character strengths? Is there a strong and potentially meaningful general factor (like the *G* factor of intelligence)? Are there meaningful metatraits (similar to Stability/Plasticity or the Big Two)? Insights into these questions can also stimulate future theorizing as to the sources and functionality/adaptivity of individual differences in character (e.g., in terms of cognitive, genetic, evolutionary, or cultural processes and correlates).

Second, it facilitates comparisons with other trait spaces and their hierarchy allowing to establish points of convergence and divergence. For example, do the global traits on the uppermost levels of the VIA trait space resemble global traits in other trait spaces, such as the Big Two metatraits or the Big Five domains in the personality trait hierarchy? From a conceptual point of view, there is reason to expect both similarities and differences between the VIA trait space and the lexical personality (Big Five and HEXACO/Big Six) trait space. This is because VIA comprises purposively selected, positively valued (i.e., evaluative) traits but is not lexically exhaustive. For the same reason, however, VIA may cover content (e.g., spirituality, humor, valor, social intelligence) that is insufficiently represented in the Big Five and HEXACO/Big Six frameworks. Therefore, the comparison of the VIA trait hierarchy and the personality trait hierarchy is instructive.

Third, establishing the trait hierarchy enables researchers to measure constructs on different levels of abstraction that best fit their specific research questions. Different levels of abstraction may be most suitable for different purposes. For example, a researcher interested in parsimonious description may prefer to measure only a few global dimensions, which is also sufficient if their lower-level dimensions show equal association patterns (Mõttus et al., 2020). By contrast, another researcher interested in more fine-grained description, prediction, or explanation may opt for a more high-dimensional model comprising all 24 character strengths or even single items, sometimes called “nuances” (Mõttus et al., 2020; see also Danner et al., 2021). By analyzing all single items of a VIA inventory, this researcher may hope to exploit the total information available and sidestep potential issues of aggregate constructs, such as their questionable causal status and multi-determined nature (i.e., a score/value on a higher-level construct can correspond to multiple configurations of its lower-level constituents; Mõttus et al., 2020). Then again, offering different aggregation levels also allows for predictive research that is aligned with the Brunswikian symmetry principle. The Brunswikian symmetry principle states that the maximum possible association between two constructs is strongest when both constructs are on the same level of abstraction (Nesselroade & McArdle, 1997; Wittmann, 1988). Accordingly, the 24 VIA character strengths—which are similar in their level of abstraction to personality facets (see Bluemke et al., 2021; McGrath et al., 2020; the tables from Bluemke et al., 2021, in the SOM on OSF)—may maximize predictive power for narrow criteria, whereas global dimensions are sufficient and may even have higher predictive power for broader criteria (e.g., Mõttus et al., 2017). Moreover, not all research contexts allow for the assessment of all 24 character strengths: Multi-theme surveys in which questionnaire space is limited may choose to assess only a few global dimensions.

Previous Factor-Analytic Studies on the VIA Character Strengths

Although understanding the hierarchical nature of a construct is thus important and indeed standard in much of individual differences research, a consensus on the VIA trait hierarchy has not yet emerged. Previous empirical studies on the factor-analytic structure of the VIA trait space predominantly used Peterson et al.'s (2005) original 240-item VIA-IS inventory to measure the 24 character strengths, while some used various short forms (the shortest consisting of 24 items) or language adaptions (e.g., Polish, Spanish, Portuguese, Croatian, Hebrew, Chinese, or Urdu) of it. Most of these studies employed exploratory factor-analytical techniques (e.g., PCA or EFA) to identify a single most useful or plausible global level in the solutions-hierarchy on which to aggregate the variance contained in the 24 character strengths. The number of factors or components that were retained varied widely across these studies: one (Noronha et al., 2015; Seibel et al., 2015; Singh & Choubisa, 2009), three (Castro Solano & Cosentino, 2018; Duan et al., 2012; McGrath, 2015; McGrath & Wallace, 2021; Redfern et al., 2014; Shryack et al., 2010), four (Anjum & Amjad, 2019; Brdar & Kashdan, 2010; Macdonald et al., 2008; Najderska & Cieciuch, 2018; Petkari & Ortiz-Tallo, 2018; Xie, 2015), or five (Azañedo et al., 2014; Höfer et al., 2019; Littman-Ovadia, 2015; Littman-Ovadia & Lavy, 2012; McGrath, 2014, 2015; Peterson et al., 2008; Peterson & Seligman, 2004; Ruch et al., 2010; Singh & Choubisa, 2010).[5]

According to Ng et al. (2017), five-dimensional solutions appear to be most common, whereas other authors have highlighted a three-dimensional solution comprising the dimensions *caring* (interpersonal strengths), *inquisitiveness* (intellectual strengths), and *self-control* (intrapersonal strengths) as most reproducible across different VIA instruments, samples, and analytical strategies (McGrath, 2015; McGrath et al., 2018; McGrath & Wallace, 2021). Thus, factor-analytic findings regarding the number and nature of global dimensions in the VIA trait space are inconsistent.

The inconsistent results of previous studies most likely stem from differences across studies in the statistical analyses, sample composition and quality, and the quality of the VIA instruments. First, most studies used open-ended, exploratory factor-analytic approaches with their results strongly depending on the specific implementation of the statistical analysis. Whereas several studies showed that their findings were robust to different extraction methods (e.g., PCA or principal axis factoring) and rotation methods (i.e., orthogonal or oblique) (McGrath, 2014, 2015; Redfern et al., 2014; Shryack et al., 2010), different factor retention criteria might have contributed to the inconsistency of results. Some studies applied the Kaiser criterion (i.e., retaining factors that have an eigenvalue greater than one), which can result in the retention of too many factors or components (Zwick & Velicer, 1986). Also, parallel analysis (PA; Horn, 1965), which aims to overcome limitations of the Kaiser criterion, may result in the retention of too many factors or components for large sample sizes (Revelle, 2019a). Therefore, it may be advisable to replicate PA results based on a large sample with a smaller subsample and to use a further method to determine factor retention alongside PA, for example Velicer's (1976) Minimum Average Partial (MAP) method. However, determining a "single best" global level solely based on factor retention criteria is incompatible with the idea of a trait hierarchy, which can encompass more than one useful global level. Yet, only very few studies (e.g., McGrath, 2015; Shryack et al., 2010) used the Bass-ackwards approach (Goldberg, 2006) to unravel the VIA solutions-hierarchy. Moreover, none of the previous factor-analytic studies reported whether results were robust to using (disaggregated) item scores instead of the 24 (aggregated) scale scores as input for the factor analyses.

Second, sample composition and quality varied widely across previous studies. Most studies were based either on student samples (Brdar & Kashdan, 2010; Duan et al., 2012; Macdonald et al., 2008; Noronha et al., 2015; Petkari & Ortiz-Tallo, 2018; Singh & Choubisa, 2009, 2010; Xie, 2015) or on convenience samples that were biased towards a specific demographic group (e.g., the highly educated or females; e.g., Azañedo et al., 2014; Castro Solano & Cosentino, 2018; Littman-Ovadia, 2015; McGrath, 2014, 2015; McGrath & Wallace, 2021; Ng et al., 2017; Peterson et al., 2008; Redfern et al., 2014; Seibel et al., 2015; Shryack et al., 2010). Such selective samples are likely to suffer from restriction of range and reduced variance, which may limit the number of relevant dimensions that can be identified. Some studies used samples of non-native speakers, who may not have understood the items correctly, or a mixture of native speakers and non-native-speakers (e.g., Ng et al., 2017; Petkari & Ortiz-Tallo, 2018; Singh & Choubisa, 2010). Only few studies (e.g., Höfer et al., 2019; Ruch et al., 2010) used large samples with $N > 1{,}000$ that were drawn at random or at least were sufficiently diverse.

Third, most studies used the original 240-item VIA-IS (Peterson et al., 2005), one of its short forms or language adaptions, or alternatively the IPIP-VIA version available on the IPIP website (Goldberg et al., 2006). Despite their merits, all these instruments have psychometric shortcomings that threaten the validity of factor-analytic studies: Several items lack content validity because they do not represent the definitional core of a strength well or are too situation-specific (Bluemke et al., 2021; McGrath & Wallace, 2021). Second, the item content of some character strengths scales is too disparate (McGrath, 2014, 2019; McGrath & Wallace, 2021), hampering unidimensionality and complicating the computation of meaningful aggregate scores (Bluemke et al., 2021; McGrath, 2014; Ng et al., 2017). Many items do not load on the assigned common factor or require several cross-loadings on other strengths (Bluemke et al., 2021; McGrath, 2014; Ng et al., 2017), which shows that these items are not pure measures of the targeted character strengths. In addition, because of an exclusive or imbalanced use of positively keyed items, scale scores based on the VIA-IS or the IPIP-VIA cannot be adequately corrected for acquiescent responding (“yeah-saying”). Acquiescence is a widespread source of bias especially in cross-cultural research (Lechner et al., 2019). If not adequately corrected for, acquiescence can bias means and covariance-based statistics of items and scale scores, including any higher-level dimensions aggregated from these scales. Finally, the different VIA-IS variants and the IPIP-VIA have been criticized for their limited cross-cultural applicability (Bluemke et al., 2021; McGrath, 2019) because they make use of idiomatic item wording, which threatens precise translatability, or ask about culture-specific behavior that is not a universal indicator for a character strength. This may have contributed to the inconsistencies in factor-analytic studies that used different language adaptions of the VIA-IS.

The Present Study

In the present study, we revisit the structure of the VIA trait space through the factor-analytic approach. Our aim is to identify useful global levels of abstraction or aggregation in the VIA trait hierarchy that allow to describe the VIA trait space with greater generality and parsimony than the 24 specific character strengths—much akin to what the global Big Two, Big Five or HEXACO/Big Six traits represent in the lexical tradition of personality research. To overcome the aforementioned methodological limitations that have contributed to the inconsistencies in previous research on the hierarchical structure of the VIA trait space, we adopted a refined methodological approach: We drew on large and heterogeneous quota samples from two countries, Germany and the UK. These samples helped to prevent restriction on range, and the two-country design allowed us to investigate the generalizability of our findings. We assessed the 24 character strengths with the German- and English-language versions of the IPIP-VIA-R inventory (Bluemke et al., 2021). IPIP-VIA-R was psychometrically refined with regard to content-validity and cross-cultural applicability, scale length, balanced item keying, essential unidimensionality, and discrimination (i.e., reduced overlap) of the 24 VIA scales. The scale scores were the focus in our main analyses. To add further rigor, we also conducted a robustness check in which we used the 96 item scores as input instead of the 24 scale scores.

To identify useful global aggregation levels above the 24 VIA character strengths, we first established the dimensionality—that is, the number of relevant dimensions (i.e., principal components) that span the VIA trait space populated by the 24 character strengths—through Horn's (1965) PA and Velicer's (1976) MAP procedure. We then unfolded the solutions-hierarchies with Goldberg's (2006) Bass-ackwards approach and examined the cross-cultural replicability of the dimensions on each level by means of component congruency analysis (Tucker, 1951).

We defined three criteria to judge whether each level in the solutions-hierarchy represented a "useful" global aggregation level. These criteria refer to the internal structure and the robustness/replicability of the higher-level dimensions on each level. They flow from, and are fully consistent with, our aim to identify *useful global levels of abstraction* above the 24 character strengths:

(1) *Interpretability*: A level in the solutions-hierarchy is "well-interpretable" if all of its dimensions are characterized by a unique set of highly-loading "marker" strengths (i.e., character strengths with a loading of $\lambda \geq .50$; for details, see Method). That is, the strengths that load highly on one dimension should not load highly on other dimensions of the same level, such that all dimensions represent the essence of a different set of strengths. This criterion ensures that all higher-level dimensions are distinct and can be meaningfully interpreted, labelled, as well as communicated.

(2) *Globality/Generality*: A level in the solutions-hierarchy is "global" if all dimensions on that level represent more abstract, general concepts that express what several of the more specific character strengths have in common. We therefore stipulate that each dimension of a useful global level of aggregation should bundle the essence of at least three character strengths. For this to be the case, we stipulate that a global dimension should comprise at least three highly-loading marker strengths ($\lambda \geq .50$).

(3) *Cross-cultural replicability*: A level in the solutions-hierarchy is "cross-culturally replicable" if the patterns of loadings show high cross-country congruency (Tucker's $\Phi \geq .90$; for details, see Method). Cross-national replication is a sign that the higher-level dimensions are robust (i.e., not a chance finding). It is also a precondition for their applicability in cross-cultural research and their status as potential human universals.

Method

Data

In each of two data collections (in 2018 and 2019), we sampled respondents in Germany and UK through a commercial online access panel provider. We pooled the non-overlapping data from both collections within each country. For both data collections, we drew a German quota sample based on gender, age, and level of education that matched German census data and a parallel sample in UK. The four initial samples comprised N = 518 and N = 509 respondents in Germany, and N = 522 and N = 524 in UK, totaling N = 1,027 in Germany and N = 1,046 in UK.

To ensure data quality, we excluded careless responders based on the Mahalanobis distance of the individual response vector from the mean sample response vector (Meade & Craig, 2012), the ipsatized variance across item scores (DeSimone & Harms, 2018), and the average response time per item (Leiner, 2019). In each sample from Germany and UK separately, respondents were flagged as careless responders if they fell within the upper 2.5% of the sample distribution of the Mahalanobis distance, or the lower 5% of the sample distribution of the ipsatized variance, or if their average response time per item was ≤ 1 second. Most estimates of the proportion of carless responders in a survey range between 5–15% (DeSimone & Harms, 2018). For example, Meade and Craig (2012) detected 10–12% careless responders in a student sample. Assuming that this can be considered as upper limit for our samples, we aimed for an exclusion rate below 10%. Using this approach, 83 respondents in Germany (8.08%) and 96 respondents in UK (9.18%) were flagged as careless responders and excluded from the analyses. Table 2.1 shows the composition of the pooled final analysis samples after exclusion of careless responders. The share of missing values on VIA item scores and scales scores calculated therefrom were negligible (six values in total across both analytical samples). The input correlation matrices of the VIA variables were based on pairwise complete cases.

Table 2.1

Sample Characteristics in Germany and UK

	Germany	UK
N	944	950
Age in years, *M* (*SD*) [*range*]	43.79 (14.92) [18–69]	44.30 (14.34) [18–69]
Proportion of women (%)	50.85	51.37
Educational level (%)		
Low	34.43	33.58
Intermediate	32.84	33.68
High	32.73	32.74

Note. Educational levels: low: no educational qualification, lower secondary leaving certificate; intermediate: intermediate school leaving certificate; high: higher education entrance qualification.

Open Science and Transparency Statements. The factor-analytic methods in our study required sample sizes large enough to ensure stable correlation matrices and patterns of loadings. By pooling data from two data collections in both Germany and UK, we ensured large absolute sample sizes and high subject-to-item ratios (approx. 40:1 in scale-based analysis and approx. 10:1 in item-based analysis; see Osborne & Costello, 2004). The samples were also four times larger than the sample sizes typically required for sample correlation matrices to stabilize according to simulation studies (Schönbrodt & Perugini, 2013).

The merged dataset including a filter variable for careless responders (which we applied to obtain our final analysis samples) can be retrieved from the project website at OSF. We also provide the codebooks of the data collections to make additional variables, which were not used in the present paper, evident.

A subset of about 50% of the data used in our present study (i.e., the first collection in 2018) constituted the cross-replication sample for the development of the IPIP-VIA-R (Bluemke et al., 2021). There it was used for the validation of each of the 24 character strength scales and computation of statistical indices for unidimensionality, reliability, construct and criterion validity, and cross-cultural measurement invariance. None of the data were previously used to analyze the hierarchical structure of the VIA trait space.

Measures

We assessed the VIA character strengths with the 96-item IPIP-VIA-R inventory (Bluemke et al., 2021) which will be published on the IPIP website at https://ipip.ori.org/. We also provide both the IPIP-VIA-R inventory and tables from Bluemke et al. (2021), which attest to the psychometric quality of the IPIP-VIA-R, on the project's OSF website. The IPIP-VIA-R is a purposeful item selection from the International Personality Item Pool (IPIP; Goldberg et al., 2006). IPIP-VIA-R measures each of the 24 VIA character strengths with a balanced-keyed set of four items (i.e., two positively and two negatively keyed) that were selected to refine content-validity (i.e., their compliance with the definition of each character strength), essential unidimensionality, discrimination (i.e., reduced overlap between scales aiming at reduced cross-loadings), and cross-cultural applicability. In Germany, the bivariate correlations between its 24 unit-weighted scale scores ranged from $r = -.16$ to $r = .61$, with an average correlation of $r = .31$. In UK, the bivariate correlations ranged from $r = -.10$ to $r = .64$, with an average correlation of $r = .36$. We estimated the composite reliability for each of the 24 unit-weighted scale scores based on a latent measurement model, taking into account the categorical nature of the response scales (Bluemke et al., 2021). The MIMIC measurement model comprised a latent character strength variable measured with four items and the ipsative mean across the 96 VIA items as an observed exogenous covariate. This covariate explained/removed the variance proportion in each item that resulted from acquiescent responding, a major source of bias especially in cross-cultural research (Lechner et al., 2019).

We refer to our reliability estimates as ω_H because of their conceptual similarity with Reise et al.'s (2013) omega hierarchical: they reflect only the common variance of the four items represented in the latent character strength variable as reliable variance, but not the acquiescence variance. The average ω_H across the 24 short-scales was .75 (*SD* = .07) and ranged between .55–.85 in Germany. In UK, average ω_H was .76 (*SD* = .05) and ranged between .62–.82. We estimated test-retest reliability (r_{tt}) based on subsamples of the analysis samples with comparable sample composition (*N* = 228 in Germany, *N* = 225 in UK) and a 2–3 weeks test-retest interval. The average r_{tt} of the 24 scales was .73 (*SD* = .06) and ranged between .59–.85 in Germany. In UK, the average r_{tt} was .67 (*SD* = .06) and ranged between .55–.79. ω_H and r_{tt} coefficients were computed in Mplus (version 8.3; Muthén & Muthén, 1998–2017). Descriptive information, ω_H, and r_{tt} for each scale in Germany and UK are provided as SOM on the OSF project website.

Statistical Analyses

All subsequently described statistical analyses were conducted in R (version 3.6.1; R Core Team, 2019). All Mplus and R code and information on R package versions can be retrieved from the OSF project website.

Our factor-analytic approach consisted of three complementary steps (described in detail below). The 24 unit-weighted scale scores representing 24 distinct character strengths served as input in all steps. As a robustness check, we re-ran all analyses using the 96 item scores as input. We summarize the results of this robustness check in the main article and report details in the SOM on the OSF project website. We purged both the 24 scale scores and the 96 item scores of acquiescence variance, although in different ways: Whereas acquiescence is roughly averaged out in scores built from balanced scales, we corrected item scores for acquiescence bias by means of ipsatizing (Billiet & McClendon, 2000; Lechner et al., 2019).

Because we do not assume the relationship between the character strengths and their higher-level dimensions necessarily to be causal (in neither direction) and the higher-level dimensions to be latent constructs or variables, we represented the higher-level dimensions as principal components, thereby treating the higher-level dimensions as weighted linear composites of the character strengths (see Bollen & Diamantopoulos, 2017; Edwards, 2011; Mõttus & Allerhand, 2018 for an in-depth discussion on the properties of PCA). Furthermore, the use of principal components is in line with most previous factor-analytic studies on the VIA trait space (e.g., McGrath, 2015; Ruch et al., 2010; Shryack et al., 2010) and the theoretical approach by Peterson and Seligman (2004) in which a non-causal relationship between the 24 character strengths and the six core virtues is assumed.

Step 1: PA/MAP. We used PA (Horn, 1965) and MAP (Velicer, 1976) in Germany and UK. PA and MAP are different methods—whose results need not always agree—to determine how many relevant dimensions (i.e., principal components) can be extracted from a correlation matrix in order to parsimoniously summarize the (co-)variation contained in the set of input variables. In PA, empirical eigenvalues are compared with random eigenvalues to determine how many components in the correlation matrix at hand “are meaningfully different from random noise” (Lang et al., 2016, p. 39). We conducted PA with the psych package (Revelle, 2019b) and used the 95th quantile (Glorfeld, 1995) of 1000 resampled or simulated data matrices to determine the random eigenvalues. Because PA is sensitive to sample size in a way that larger samples might result in an over-extraction of components (Revelle, 2019a), we conducted the PA in both countries based on both the full analysis sample and a random subsample of *N* = 500 each to check the robustness of the full-sample based results.

We conducted MAP with the EFA.dimensions package (O’Connor, 2020). Whereas PA takes the total variance of the input variables into account, MAP focuses on their common variance. MAP works by partialling out an increasing number of components from the variables' correlation matrix and stopping when the average squared partial correlation of the off-diagonals is minimized (O'Connor, 2000; Velicer et al., 2000). We used the revised MAP criterion in which the partial correlations are taken to the fourth (instead of second) power (Velicer et al., 2000).

Note that we used PA and MAP as helpful guidance but not as key criteria. Although PA and MAP are highly informative as to how many strong and relevant dimensions can be extracted from a correlation matrix, their results do not dictate a single "correct" number of dimensions to retain (or, equivalently, single "correct" level of the Bass-ackwards analysis; Goldberg, 2006). For example, even in cases in which PA/MAP suggest the extraction of, say, 4 dimensions, a 5- or even 6-dimensional solution might still be useful, depending on the criteria of usefulness. Therefore, when scrutinizing global levels of a trait hierarchy, researchers often do not stick strictly to PA/MAP results but give more weight to other criteria (e.g., Lang et al., 2016)—as we do in the present study. None of our previously defined criteria of what constitutes "useful" global levels (interpretability, globality/generality, cross-cultural replicability) depends solely on PA/MAP, although interpretability and globality/generality are of course not fully independent of PA/MAP.

Step 2: Bass-Ackwards Analysis. To unfold the solutions-hierarchy of the VIA strengths, we conducted Bass-ackwards analyses (Goldberg, 2006) both in Germany and in UK. The Bass-ackwards approach is a simple exploratory procedure to investigate the hierarchical structure of a set of variables. Using the unit-weighted scale scores for each of the 24 character strengths as input, we conducted PCAs extracting an increasing number of components—first one component, then two components, and so forth. We then computed correlations between components of adjacent levels. For our Bass-ackwards analyses, we used the R code provided by Waller (2007).[6] We modified Waller's function to apply to obliquely rotated principal components (Promax rotation, m = 4).

To facilitate the interpretation and evaluation of each dimension, we classified the character strengths depending on the size of their loadings: We classified strengths that loaded with $\lambda \geq .50$ on a higher-level dimension as its *marker strengths*, because these strengths play a major role in defining the substantial meaning of the dimension. Likewise, we classified strengths that loaded with $.30 \leq |\lambda| < .50$ on a higher-level dimension as *co-defining strengths*, because these strengths play a secondary/minor but non-negligible role in defining the substantial meaning of the dimension. We disregarded strengths that loaded with $|\lambda| < .30$ on a dimension.[7]

Based on the resulting solutions-hierarchy, we identified those levels whose dimensions (i.e., components) met our interpretability and generality/globality criterion. We examined and reported Levels I–VIII of the solutions-hierarchy, but not lower levels, because—as usually in Bass-ackwards analyses—our focus in this study was on identifying *global* dimensions. Beyond Level VIII, it is numerically impossible that the dimensions of a level could comprise a unique set of minimum three highly-loading marker strengths and thus meet our interpretability and globality/generality criterion.

Step 3: Component Congruency. We identified those levels of the solutions-hierarchy that comprised the most similar and thus most replicable dimensions across Germany and UK (our third criterion). We first rotated the principal component solution at each level to maximum similarity between countries, using the component loading matrix (pattern matrix) obtained in Germany as the target matrix. For this target rotation, we used Jennrich's (2002) gradient projection rotation optimization algorithm as implemented in the GPArotation package (Bernaards & Jennrich, 2005).[8] We then computed Tucker's Phi (Tucker, 1951) to gauge the congruency across countries of the target-rotated components on the same hierarchical level. The size of Tucker's Phi is independent of the mean absolute size of component loadings and expresses similarity in terms of profile similarity but not in terms of a similar amount of explained variance (Lorenzo-Seva & ten Berge, 2006). Following

Lorenzo-Seva and ten Berge (2006) and Jensen (1998), we interpreted values of Tucker's Phi of $\Phi \geq .95$ as essentially equivalent, $\Phi \geq .90$ as highly similar, and $\Phi \geq .85$ as fairly similar.

Results

PA & MAP

Figure 2.1 shows scree plots obtained from the PAs. In both countries, a strong first component emerged (eigenvalue > 8 in Germany and > 9 in UK, respectively) in both the full

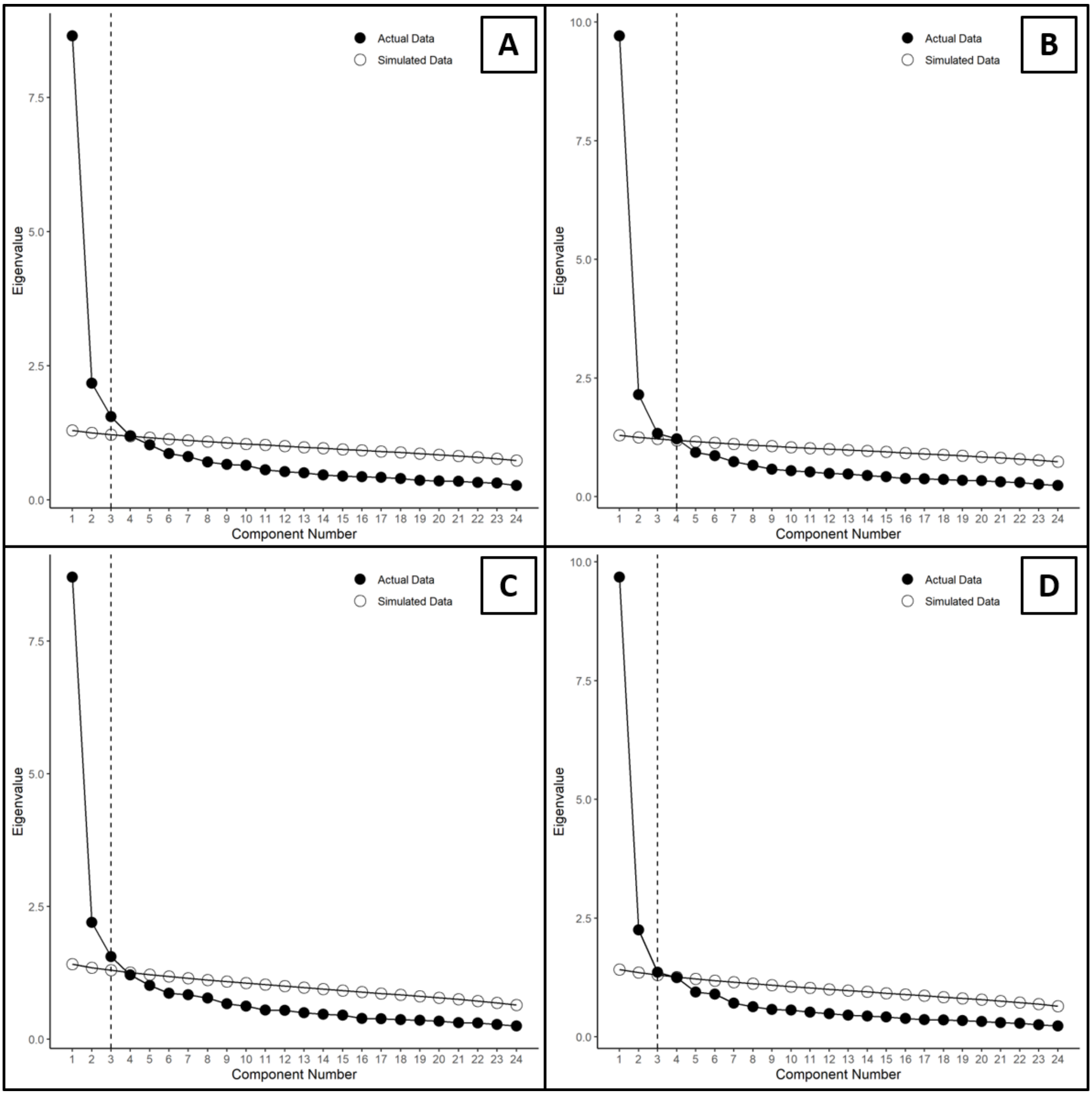


Figure 2.1. Results from PAs. The plots show the eigenvalues of the first to 24th principal component extracted from actual and simulated or resampled data. Panel A: full analysis sample Germany (*N* = 944). Panel B: full analysis sample UK (*N* = 950). Panel C: random subsample Germany (*N* = 500). Panel D: random subsample UK (*N* = 500).

samples (panels A & B) and the random subsamples (panels C & D). The second component already had a much smaller eigenvalue (< 2.5 in all samples). Overall, PA suggested retaining three components—somewhat more unambiguously in Germany (panels A & C) than in UK (panels B & D). Likewise, MAP suggested retaining three components in both countries. Based on the full analysis samples, these three (unrotated) components explained 36.05%, 9.05%, and 6.46% of the total variance (i.e., 51.56% combined) in Germany and 40.45%, 8.96%, and 5.56% (i.e., 54.97% combined) in UK, respectively.

These results show a strong saturation of the first component and suggest that three global dimensions suffice to capture the essence of the 24 VIA character strengths as measured with the IPIP-VIA-R. A three-component solution was sufficient to recover more than half of the total variation in the 24 character strengths in both countries.

Bass-Ackwards Analysis

Figures 2.2 and 2.3 show the (truncated) results from the Bass-ackwards analyses in Germany and UK. Each figure depicts Levels I−VIII of the solutions-hierarchy with the rectangles representing obliquely rotated principal components and the coefficients along the paths expressing correlations of components of adjacent levels. To evaluate the components according to our criteria of interpretability and globality/generality, the rectangles contain lists of their *marker strengths* (i.e., character strengths with loadings of $\lambda \geq .50$ on that component; top rows) and *co-defining strengths* (i.e., character strengths with loadings ranging of $.30 \leq |\lambda| < .50$; bottom rows) in descending order. Arabic numerals denote cross-culturally corresponding components at a hierarchical level across Figures 2.2 and 2.3. For Germany, the Arabic numerals indicate the order of extraction within each solution or level. For UK, at some levels the components were extracted in a different order than for Germany. (We provide the detailed outputs of the Bass-ackwards analyses including loading matrices, between- and within-level correlations of the Levels I–XXIV on the project website on OSF.)

APP Appreciation of beauty
CAP Capacity for love
CIT Citizenship/Teamwork
CUR Curiosity
EQU Equity
FOR Forgiveness
GRA Gratitude
HOP Hope
HUM Humor
IND Industriousness/Perseveranc
INT Integrity
JUD Judgement
KIN Kindness
LEA Leadership
LOV Love of learning
MOD Modesty/Humility
ORI Originality
PER Perspective
PRU Prudence
SEL Self-regulation
SOC Social intelligence
SPI Spirituality/Religiousness
VAL Valor/Bravery
ZES Zest

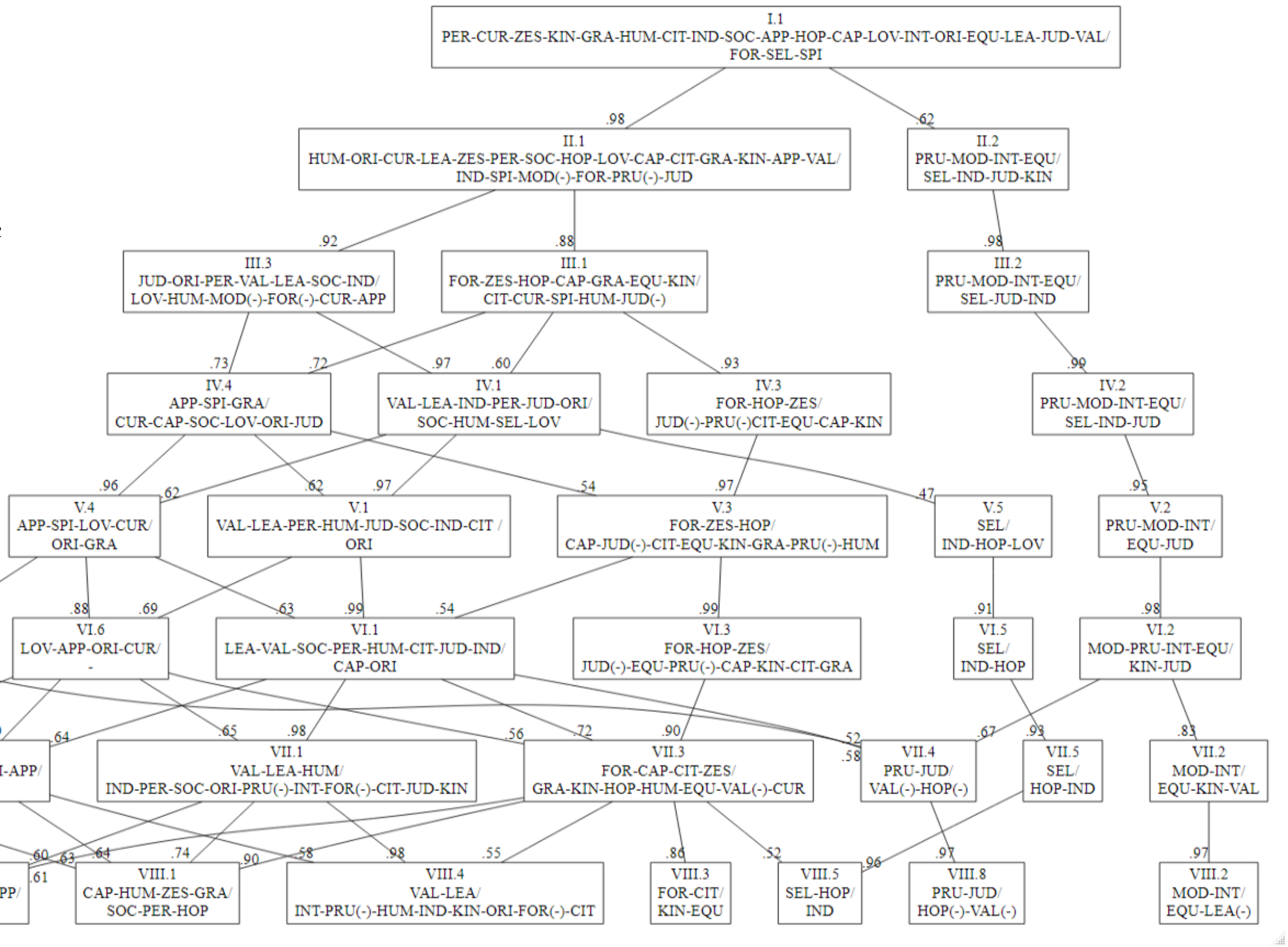


Figure 2.2. Levels I–VIII of VIA solutions-hierarchy in Germany. Boxes represent obliquely rotated principal components. Roman numerals denote hierarchical levels. Arabic numerals denote the order in which components of a hierarchical level were extracted. Character strengths are depicted in descending order of (absolute) loading size with marker strengths ($\lambda \geq .50$) in top row and co-defining strengths ($.30 \leq |\lambda| < .50$) in bottom row. (-) denotes inversely loading character strengths. Within-level correlations: $r_{\text{II.1/II.2}} = .46$, $r_{\text{III.1/III.2}} = .33$, $r_{\text{III.1/III.3}} = .62$, $r_{\text{III.2/III.3}} = .31$, $.20 \leq r_{\text{IV}} \leq .56$, $.06 \leq r_{\text{V}} \leq .61$, $-.05 \leq r_{\text{VI}} \leq .67$, $-.03 \leq r_{\text{VII}} \leq .61$, $-.14 \leq r_{\text{VIII}} \leq .66$.

APP Appreciation of beauty
CAP Capacity for love
CIT Citizenship/Teamwork
CUR Curiosity
EQU Equity
FOR Forgiveness
GRA Gratitude
HOP Hope
HUM Humor
IND Industriousness/Perseveranc
INT Integrity
JUD Judgement
KIN Kindness
LEA Leadership
LOV Love of learning
MOD Modesty/Humility
ORI Originality
PER Perspective
PRU Prudence
SEL Self-regulation
SOC Social intelligence
SPI Spirituality/Religiousness
VAL Valor/Bravery
ZES Zest

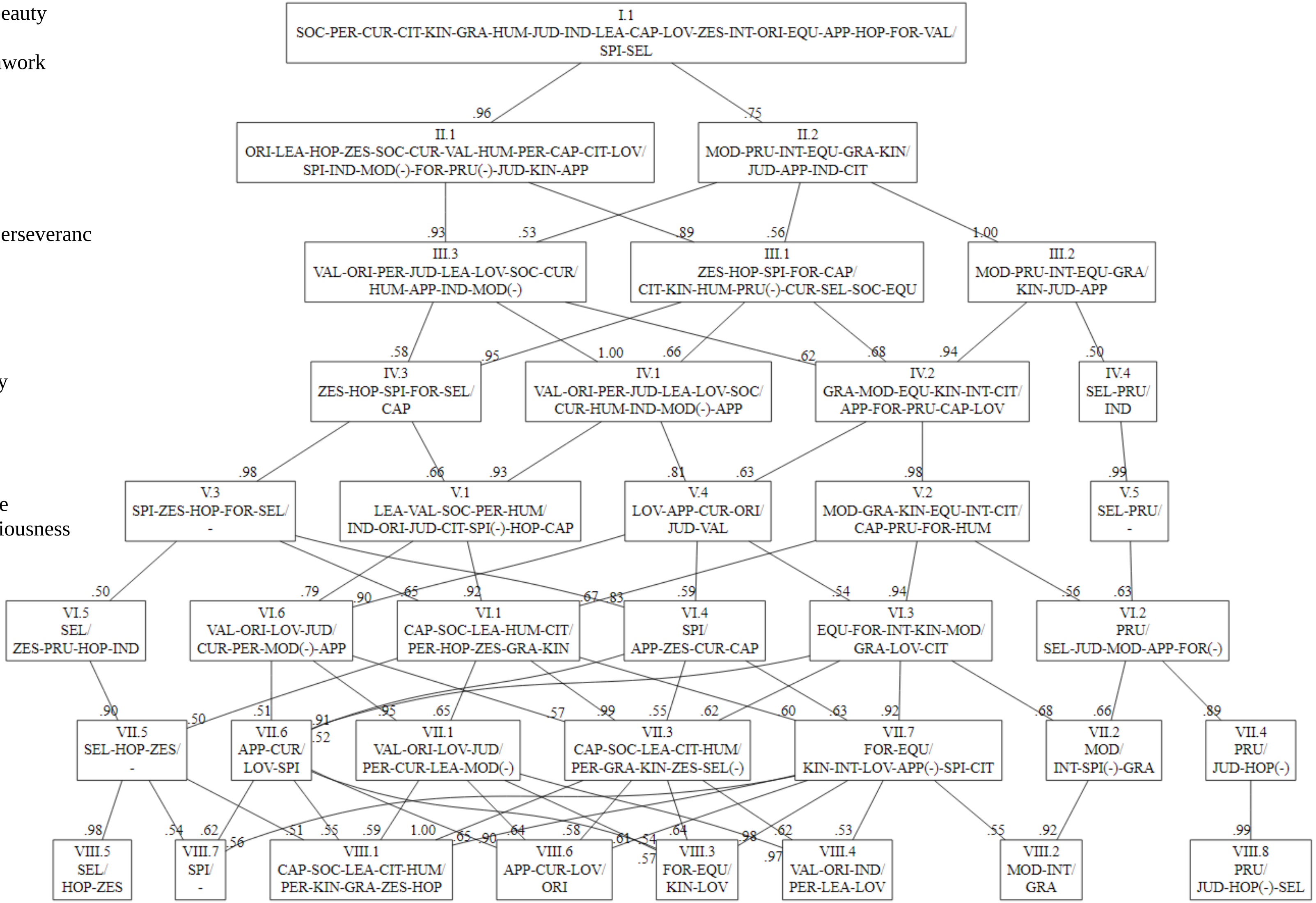


Figure 2.3. Levels I–VIII of VIA solutions-hierarchy in UK. Boxes represent obliquely rotated principal components. Roman numerals denote hierarchical levels. Arabic numerals denote a component's counterpart in Germany. Character strengths are depicted in descending order of (absolute) loading size with marker strengths ($\lambda \geq .50$) in top row and co-defining strengths ($.30 \leq |\lambda| < .50$) in bottom row. (-) denotes inversely loading character strengths. Within-level correlations: $r_{\mathrm{II.1/II.2}} = .53$, $r_{\mathrm{III.1/III.2}} = .49$, $r_{\mathrm{III.1/III.3}} = .68$, $r_{\mathrm{III.2/III.3}} = .50$, $-.17 \leq r_{\mathrm{IV}} \leq .58$, $-.17 \leq r_{\mathrm{V}} \leq .55$, $-.27 \leq r_{\mathrm{VI}} \leq .61$, $-.07 \leq r_{\mathrm{VII}} \leq .66$, $-.11 \leq r_{\mathrm{VIII}} \leq .65$.

Level I. The strong first (and sole) component at Level I expressed a manifold of positively loading marker strengths in both Germany and UK. Perspective, curiosity, kindness, gratitude, and humor were among the highest-loading marker strengths in both countries. In Germany, 19 of the 24 character strengths had loadings of $\lambda \geq .50$ on the first component, whereas that number was 20 in the UK. In addition to the large number of highly-loading marker strengths, in both countries, self-regulation and spirituality were among the moderately loading co-defining strengths ($.30 \leq |\lambda| < .50$).

The large number of character strengths with high or moderate loadings again indicated a high saturation of the first component. Only two of the 24 strengths—modesty/humility and prudence—showed negligible loadings on the first component. This suggests that these two character strengths are different from most other character strengths and do not share their common core.

Level II. In both countries, component II.1 (subsequently, we refer to components only by their numerical code) was highly correlated with the sole component at Level I. Originality, leadership, and zest were among the highest-loading marker strengths of this component in both Germany and UK. In contrast, yet consistent across countries, II.2 was characterized primarily by those strengths that appeared most distinct in the one-component solution, namely prudence and modesty/humility, now complemented by integrity and equity. A substantial correlation between II.2 and the sole component at Level I emerged in both countries, albeit lower than that of II.1.

Level III. At the third level of the solutions-hierarchy, II.2 re-appeared as III.2, as evidenced by the high correlations across levels between the two components in both countries. Marker strengths of III.2 in both Germany and UK were prudence, modesty/humility, integrity, and equity.

III.1 and III.3 split from II.1. Their respective marker strengths were essentially the same across countries: Forgiveness, zest, hope, and capacity for love were marker strengths in both countries for III.1, whereas judgement, originality, perspective, valor, leadership, and social intelligence were marker strengths in both countries for III.3.

Thus, all three dimensions at Level III had largely a unique set of highly loading marker strengths, and these marker strengths were essentially the same in both countries. Given the distinctness and number of marker strengths of each of the three factors, Level III can be judged favorably against our criteria of interpretability and globality/generality in both countries. (In addition to the Figures 2.2 and 2.3, we provide the loading patterns of Level III in Germany and UK including full and shortened labels of the character strengths scales as SOM on OSF.)

Levels IV−VIII. From the fourth level onward, the globality/generality criterion was no longer met in both countries. In UK, IV.4 appeared as a first “splinter” component marked by only two character strengths, namely self-regulation and prudence. In Germany, V.5 was exclusively marked by self-regulation.

On Level VI, two components in Germany and three components in UK had only one highly-loading marker strength and were therefore of insufficient globality/generality. Self-regulation and spirituality, which loaded only moderately on the sole component at Level I in both countries, splintered off and dominated separate components at Level VI in both Germany and UK. Prudence dominated another component in UK. Of note, the six dimensions at Level VI did not resemble the six core virtues proposed by Peterson and Seligman’s (2004) theoretical classification.

At Levels VII and VIII, on which a component structure with three marker strengths per component is still feasible in principle, several components were characterized by only one or two marker strengths in both countries, again undermining the globality/generality criterion. The majority of components at those levels were characterized by only few marker strengths and co-defining strengths.

In general, beyond Level III, the globality/generality of dimensions within a hierarchical level varied and decreased strongly: Whereas the global dimensions observed at Level III propagated through subjacent levels of the solutions-hierarchy largely unchanged, the additionally extracted components often were narrow being characterized by only one or two marker strengths. It is a pattern to be expected in the Bass-ackwards approach that global dimensions from higher levels reappear at (or propagate through) lower levels, whereas additionally extracted components are often narrow “splinter” components (see Goldberg, 2006; Shryack et al., 2010). Observing this pattern already from Level IV on is in line with the scree plots and PA results in Figure 2.1, which suggested that only three strong higher-level dimensions span the VIA trait space (as measured with the IPIP-VIA-R).

Cross-Country Congruency Analyses

Table 2.2 shows the congruency coefficients (Tucker’s Φ) between the components in Germany and their target-rotated counterparts in UK at Levels I–VIII of the solutions-hierarchy. At Levels I−III, all components showed congruencies of $\Phi \geq .90$ or even $\Phi \geq .95$. This indicates that the components are of high similarity or even essential equivalence across countries, which implies that their structure and meaning were replicable and almost identical across countries.

All of the subjacent Levels IV–VIII contained one or more components with congruencies of $\Phi < .90$, indicating a lower degree of similarity of the loading patterns across countries. Among these levels, Level VI stands out as relatively good, with five of its components showing $\Phi > .90$ and only VI.4 showing $\Phi = .86$, which would still be considered “fairly similar” (Lorenzo-Seva & ten Berge, 2006). Nevertheless, these results suggest that only Levels I−III, but not subjacent levels, exclusively comprised higher-level dimensions that were highly similar across Germany and UK. Consequently, Levels I–III but not subjacent levels unambiguously met our criterion of cross-national replicability.

Table 2.2
Tucker's Phi Coefficients Indicating the Congruency Across Countries of the Components at each of the Levels I−VIII

	C1	C2	C3	C4	C5	C6	C7	C8
Level I	1	-	-	-	-	-	-	-
Level II	.99	.94	-	-	-	-	-	-
Level III	.94	.93	.96	-	-	-	-	-
Level IV	.95	.96	.88	.71	-	-	-	-
Level V	.98	.95	.94	.89	.74	-	-	-
Level VI	.97	.96	.94	.86	.93	.92	-	-
Level VII	.89	.87	.88	.95	.93	.90	.60	-
Level VIII	.87	.93	.93	.93	.94	.95	.76	.97

Note. Tucker's Phi coefficients were computed after target rotation of UK components towards their corresponding components in Germany. The order of components C1−C8 corresponds to the order of extraction in Germany. Values of Φ ≥ .95 indicate that the patterns of loadings of a component are "essentially equivalent" in both countries, Φ ≥ .90 indicate they are "highly similar", and Φ ≥ .85 indicate they are still only "fairly similar" (Jensen, 1998; Lorenzo-Seva & ten Berge, 2006).

According to our criteria, we identified Levels I–III (thereof foremost Level III) *as useful global levels of abstraction* above the 24 character strengths. These levels comprise well-interpretable and cross-nationally replicable dimensions that abstract the common core from several character strengths in different ways. The dimensions on these levels likely represent global character or personality constructs that (1) can contribute to theory-building in character research and (2) allow for a parsimonious and more global assessment of individual differences in human character. Levels IV–VIII and subjacent levels, by contrast, did not qualify as "useful" global levels according to our criteria. This of course does not preclude that these more fine-grained levels are of theoretical and practical value. For example, the narrow dimensions they contain may improve prediction of specific outcomes compared to global dimensions. However, according to our criteria, these levels clearly do not constitute useful *global* levels in the sense that Levels I–III do.

Robustness Checks Using the 96 Items as Input

The results presented thus far were based on the 24 character strengths' scale scores, in line with previous studies and consistent with our aim to identify useful aggregation levels that summarize the 24 well-established character strengths. However, as a robustness check, it may be instructive to test whether the levels we identified as "useful" global levels (i.e., the Levels I–III and of these especially Level III) also emerge when using the 96 disaggregated items of the IPIP-VIA-R instead of the 24 aggregated scale scores as input. We therefore re-ran all analyses using the 96 ipsatized item scores. Although the item scores contain additional variance compared to the variance that is present in the aggregated scale scores (i.e., more specific trait variance capturing substance beyond the character strengths as well as more specific method variance), this additional variance is unlikely to be retained at high levels of aggregation. Accordingly, we expected to replicate the global dimensions from the scale-based analyses with item-based analyses. We report and discuss the results in detail in the SOM on OSF.

As to be expected from the additional variance contained in the item scores, PA and MAP performed on the 96 item scores suggested to extract 7–10 components, and hence more than in the scale-based analyses. However, a closer look revealed much greater agreement between the item-based and scale-based analyses than PA and MAP suggest. Specifically, judging the item-based results against our three criteria of what constitutes useful global levels, we found that (1) the *globality/generality* criterion only held from Levels I–III; (2) the dimensions on Levels I–III—but not subjacent levels—fulfilled the *interpretability* criterion with the item loading-based marker strengths (see below/SOM on OSF) largely corresponding to the scale-based marker strengths. Moreover, (3) item-based congruency analyses confirmed that only Level I–III reflected *cross-culturally replicable* higher-level dimensions throughout, whereas subsequent levels contained at least one component that fell short of the $\Phi \geq .85$ cut-off indicating fairly similar components (we outline in the SOM why item-based and scale-based congruency analyses are more comparable when applying $\Phi \geq .85$ in the item-based analysis and $\Phi \geq .90$ in the scale-based analysis, respectively).

These conclusions held irrespective of whether we interpreted the item-based findings in the light of the 24 character strengths or took a truly bottom-up interpretation approach that does not presume that the 24 character strengths exist. When interpreting the solutions-hierarchy in light of the 24 character strengths we looked at the sum of the absolute item loadings of each character strength to classify it as co-defining or marker strength of each dimension (more details are provided in the SOM on OSF). In the truly bottom-up interpretation approach, we looked at the item-based solutions-hierarchy by counting the single items with $|\lambda| \geq .50$ and $|\lambda| \geq .30$, respectively, ignoring which character strength they belonged to (i.e., instead of adding up the absolute loadings of the four items building a character strength scale).

In sum, in the item-based analyses, too, Level III of the solutions-hierarchy proved most appropriate to describe the VIA trait space parsimoniously and more generally with higher-level dimensions. By contrast, all levels below Level III must be discarded based on our criteria of interpretability, globality/generality, and cross-cultural replicability. Thus, despite some differences, the scale-based and item-based analyses led to the same overall conclusions.

Discussion

There is an ongoing debate about useful global levels of abstraction at which to best summarize the 24 VIA character strengths described by Peterson and Seligman (2004). In contrast to most major individual difference constructs (e.g., personality traits in the lexical tradition or intelligence), no trait hierarchy for VIA has yet been established, and no consensus has been reached as to the number and nature of higher-level dimensions that might lend themselves as useful aggregates of the 24 character strengths. In the present study, we unraveled the hierarchical structure of the VIA trait space through a factor-analytic approach. Our aim was to identify useful levels of abstraction on which to summarize the variation in the 24 VIA character strengths that are (1) well-interpretable, (2) global/general, and (3) cross-culturally replicable.

Converging Evidence for Level III as the Most Useful Global Level of Aggregation

When jointly considering our three criteria, Level III of the solutions-hierarchy clearly emerged as the most useful level for aggregating the 24 character strengths. PA and MAP agreed in suggesting that three components are enough to parsimoniously summarize the (co-)variation contained in the 24 character strengths. The first three components jointly accounted for a large proportion of the total variation in the 24 character strengths, of more than 50% in both countries in particular. Even more important, the solutions-hierarchies unfolded through the Bass-ackwards approach indicated that only Levels I–III, but not subjacent levels, consisted of exclusively global dimensions that each summarized what at least three highly-loading character strengths have in common. Moreover, component congruency analyses interpreted based on conventional cut-offs for Tucker's Φ showed that only Levels I–III, but not subjacent levels, are populated exclusively by highly similar or essentially equivalent dimensions across Germany and the UK. (Note that cut-offs are to some extent arbitrary. Applying $\Phi \geq .85$ as cut-off indicating fairly similar components would suggest considering Level VI of the scale-based solutions-hierarchy cross-nationally replicable, too.)

A closer look at the component structure of Level III revealed that all 24 character strengths appeared either as marker strengths ($\lambda \geq .50$) or as co-defining strengths ($.30 \leq |\lambda| < .50$) in both Germany and UK. In other words, none of the character strengths fell outside the trait space covered by the three dimensions. Furthermore, none of the character strengths functioned as marker strength for more than one component (the only exception was a cross-loading of equity). Consequently, as required by the interpretability criterion, each of the three dimensions was characterized by a unique set of highly-loading marker strengths which also were largely the same across countries (4–6 shared marker strengths per dimension). Cross-loadings, by contrast, were only present in co-defining strengths.

Hence, each dimension at Level III was readily interpretable based on its unique set of marker strengths that were shared across countries. However, naming global dimensions is generally challenging, as the chosen labels must be more inclusive than the constituent character strengths and ideally capture all their connotations. Further complicating the task, we were careful to avoid terms already used in Peterson and Seligman’s (2004) classification to avoid confusion. The current labels are therefore provisional and will likely be replaced by more suitable ones in the future. The provisional labels of the three dimensions are *positivity* (III.1), *dependability* (III.2), and *mastery* (III.3). *Positivity* is characterized by forgiveness, zest, hope, and capacity for love, thus describing a cheerful, optimistic, and forbearing reconciliator. *Dependability* is characterized by prudence, modesty/humility, integrity, and equity, thus describing a reliable, trustworthy, and caring “tower of strengths”. *Mastery* is characterized by judgement, originality, perspective, valor, leadership, and social intelligence, thus describing a wise and ingenious leader.

Although each dimension on Level III was identified by a unique set of highly-loading marker strengths, it should be noted that the dimensions were correlated because we opted for oblique rotation. In Germany, the correlations between positivity and dependability and between dependability and mastery were both $r \approx .30$. In UK, these two correlations were both $r \approx .50$. The correlation between positivity and mastery was the highest one in both countries ($r = .62$ in Germany, $r = .68$ in UK). In this context, it is noticeable that positivity and mastery both emerged from the same higher-level component (II.1) and maintained high correlations with their ancestor at Level II. This suggests that positivity and mastery themselves are related aspects of a more abstract concept.

Our conclusion that three dimensions are appropriate aggregate representations of the 24 character strengths (as measured with the IPIP-VIA-R) concurs with some previous findings on the factor-analytic hierarchical structure of VIA (Castro Solano & Cosentino, 2018; Duan et al., 2012; McGrath, 2015; McGrath et al., 2018; McGrath & Wallace, 2021; Redfern et al., 2014; Shryack et al., 2010). Together with these previous studies, our findings suggest that the VIA trait space (spanned by 3 global domains) is somewhat narrower than the lexical personality trait space (spanned by 5–6 global domains). This is not surprising, given that the VIA character strengths are, by design, not lexically exhaustive but represent a purposive theoretical selection of positive, morally valued traits that can contribute to a fulfilled life and are universally valued across cultures and time. Moreover, it should be noted that even though three global dimensions may suffice to capture the essence of the 24 VIA character strengths, these character strengths may still contain some content that is not well represented in, say the Big Five or HEXACO/Big Six (e.g., spirituality, humor, valor, or social intelligence; see correlations with Big Five facets in Bluemke et al., 2021; the tables from Bluemke et al., 2021, in the SOM on OSF; see also McGrath et al., 2020).

It should also be noted that even though we arrived at the same *number* of dimensions as these previous studies, the *nature* (or content) of the present three-dimensional solution differs from previous ones. Specifically, previous work has interpreted the three dimensions as representing different “targets of virtuous action” (McGrath, 2015, p. 418): the self (bundling intellectual strengths, labeled “Inquisitiveness”), others (bundling interpersonal strengths, labeled “Caring”), and the physical world (bundling intrapersonal strengths for dealing effectively upon the environment, labeled “Self-Control”) (McGrath, 2015; McGrath et al., 2018). By contrast, our results are better aligned with the logic of a factor-analytic trait hierarchy: The *specific* positively valued traits embodied by the 24 character strengths are aggregated to three *global/general* positive traits that play a role across different targets in both social and non-social contexts. For example, the character strength of perspective is not only relevant for *mastering* one's own life but is also helpful to others who seek advice from an individual who scores high on perspective.

We attribute the different nature of previously identified three-dimensional solutions to the psychometric shortcomings of previously used VIA instruments. Several items of previously used VIA instruments had questionable content validity, barely controlled acquiescent response style, or were otherwise insufficiently distinct measures of their targeted character strength, all of which resulted in insufficiently nuanced scales with substantial item cross-loadings (e.g., Ng et al., 2017). Using VIA scales that did not discriminate sufficiently well between character strengths, the character strengths may have previously formed higher-level structures based on rather crude communality among strength scales (e.g., most character strengths with emphasis on any kind of interpersonal context bundled up as the so-called "Caring"-dimension). By contrast, the IPIP-VIA-R more likely captures the character strengths in a fine-grained manner with content-valid items in essentially unidimensional (i.e., factor-pure) scales, allowing a sound higher-level structure to be revealed.

A Closer Look at the Other Levels of the Solutions-Hierarchy

Even though Level III proved to be the most useful level of abstraction above the 24 VIA strengths according to our criteria, Level II also fulfilled our three criteria and may present another useful global aggregation level for the VIA character strengths. A closer look at Level II showed that dimensions II.1 and II.2 resemble metatraits in other areas of individual difference research. In particular, they had some similarities with the two personality metatraits of *Plasticity* and *Stability* (DeYoung, 2006; DeYoung et al., 2002) and—to a greater extent—with the Big Two of *Dynamism* and *Social Self-Regulation* (SSR; Saucier et al., 2014; Thalmayer & Saucier, 2014). These metatraits, in turn, have commonalities with the two well-known axes of interpersonal perception—agency and communion (Bakan, 1966).

Specifically, dimension II.1 was characterized by the strengths of originality (i.e., having original ideas, coming up with innovative and productive ways to do things), leadership (i.e., organizing collective success of a group and fostering good working relationships among members), and zest (i.e., approaching activities with physical and mental vitality, feeling alive and energetic). As such, it was somewhat reminiscent of *Plasticity* (i.e., how to successfully integrate new information and flexibly respond to changes, as expressed by the shared variation of Extraversion and Openness) and to a greater extent reminiscent of *Dynamism* (i.e., activity, potency, ascendancy, being skillful and comfortable in social situations). II.2 was characterized by the strengths of prudence (i.e., being smart and careful about choices in the interest of avoiding undue risks and preventing regrets), modesty/humility (i.e., being modest (though realistic) about one's abilities and weaknesses, valuing contributions by others), integrity (i.e., being truthful to others and to oneself, trustworthy, and acting with moral integrity), and equity (i.e., treating people equally, with respect and in a fair and unbiased manner). As such, it was somewhat reminiscent of *Stability* (i.e., maintaining psychosocial stability, as expressed by the shared variation of Agreeableness, Conscientiousness, and Emotional Stability) and to a greater extent reminiscent of *SSR* (i.e., propriety, socialization, community, solidarity, respect, compliance).

It was evident from the high between-level correlations and the overlapping marker strengths, that II.1 (reminiscent of *Dynamism* and *Plasticity*) is largely retained in *mastery* at Level III, whereas II.2 (reminiscent of *SSR* and *Stability*) is basically retained in *dependability* at Level III. Furthermore, *positivity* at Level III captured "positive affectivity" as one specific aspect of Plasticity from II.1 (DeYoung et al., 2002). Beyond that, positivity at Level III strongly resembled "positivity" as conceptualized by Caprara et al. (2012): as confidence in other people as well as a positive perception of the self, one's life and one's future. Thus, Level III of the VIA solutions-hierarchy resembles different personality metatraits (DeYoung, 2006; Saucier et al., 2014) and additionally allows measuring a third—related but distinct—metatrait: positivity. Capturing something reminiscent of Dynamism/Plasticity and SSR/Stability, Level II also offers potentially viable higher-level dimensions for an even more parsimonious aggregation of the 24 character strengths. However, compared to Level III, Level II aggregates the 24 character strengths unevenly distributed over the components (i.e., 12 (II.1) vs. four (II.2) marker strengths shared across countries), less differentiated, and with the loss of some viable information (i.e., forgiveness as one core aspect of positivity only features among the co-defining strengths).

In contrast to some previous studies that advanced four- or five-dimensional solutions (e.g., Brdar & Kashdan, 2010; Höfer et al., 2019; Macdonald et al., 2008; Peterson et al., 2008; Singh & Choubisa, 2010) or that reproduced Peterson and Seligman’s (2004) six core virtues through highly liberal approaches (Ng et al., 2017; Vanhove et al., 2016), the utility of these levels of the VIA solutions-hierarchy (as global levels) appears questionable according to our criteria and results: At Levels IV–VIII, the dimensions within each level varied widely in their degree of cross-cultural replicability and globality/generality, with some dimensions being almost as specific as a single character strength. Furthermore, Level VI did not recover Peterson and Seligman’s (2004) six theoretically derived core virtues. For example, in Germany, VI.4 and VI.5 each featured only a single marker strength (spirituality and self-regulation, respectively) and VI.6 captured love of learning, appreciation of beauty, originality, and curiosity. To reiterate, we did not expect the factor-analytic solutions-hierarchy, which is based on observed correlations as they occur “in nature”, to coincide with Peterson and Seligman’s (2004) theoretical, Linnaean-type classification of strengths and virtues, which is based on conceptual similarity and which the authors labeled as preliminary, urging the reader “not to be too concerned about the details of how [they] classified the 24 strengths under the six virtues” (p. 31).

Might Socially Desirable Responding Explain the Strong Component Saturation?

Whenever self-rated character/personality traits are concerned, some remarks on the issue of socially desirable responding (SDR) are in order. By their very nature as positive qualities that can contribute to the well-being of the individual and others, the 24 VIA character strengths are socially desirable *by definition*. Especially in view of the high saturation of the first component, this raises the question of whether component I.1 represents a meaningful general factor (i.e., good character manifesting in a wide range of behavioral tendencies that are valued by most people) or mainly represents SDR. More generally, it directs attention to the extent to which SDR drives any correlations between the character strengths—and hence the loadings and component correlations in the solutions-hierarchy. We provide a detailed evaluation of the role that SDR plays in our findings in the SOM on OSF. We discuss six different aspects related to SDR: (1) Steps taken to lower SDR during the construction process of the IPIP-VIA-R; (2) the low-stakes survey context; (3) the only moderate skewness of the items and scale scores; (4) previous evidence suggesting that SDR is only a weak source of variance in VIA questionnaires; and (5) evidence that SDR in a survey (e.g., on VIA) is not only a response style but may partly translate to socially desirable conduct in real life; as well as (6) the differentiated and complex pattern of loadings and correlations between the components on the Levels I–III (and the on average moderate bivariate correlations between the 24 strengths scales—r = .31 in Germany and r = .36 in UK—from which they result). Considering all of these aspects leads us to be optimistic that SDR is unlikely to be a strong, let alone the main, driver of the high saturation of component I.1, the loadings and component correlations in the subjacent levels of the solutions-hierarchy, and the pattern of correlations between the 24 character strengths more generally. Instead, there is reason to assume that the loadings and correlations are driven primarily by content, not style. That said, future investigations may want to further expand on the issue of SDR in the higher-level dimensions we identified.

Limitations and Directions for Future Research on Character Traits

Still, our study has some limitations that future studies should address. Our results are based on a single instrument, the IPIP-VIA-R. Even though it is safe to conclude that this inventory allows for an improved assessment of the 24 VIA strengths compared to its ancestor, and probably most other available VIA instruments, no single instrument is perfect. In order to conclusively establish useful global levels of the VIA trait hierarchy, future research should replicate our findings using other VIA instruments—ideally new instruments that provide good content validity, approximation to unidimensionality and cross-cultural applicability. At the same time, we note that there are already studies based on other VIA instruments that agree with ours in highlighting Level III as useful global level in the VIA trait hierarchy, for example, a recent study of McGrath and Wallace (2021) who employed a revised version of the VIA-IS, the 192-item VIA-IS-R measuring each character strength with eight items.

Although we advanced over previous studies by using large and diverse samples from two countries, future studies would benefit from using true random samples from the general population and especially from investigating a larger set of cultures and languages. Doing so would help establish whether the promising three-dimensional solution is cross-culturally replicable beyond the two WEIRD (Western, educated, industrialized, rich, and democratic) countries investigated here, Germany and UK. For this to be a feasible goal, high-quality translations and proper cultural adaptations of IPIP-VIA-R or other VIA instruments need to be developed. It will also be of particular importance to replicate our results with informant/observer reports to see whether such ratings conform to the same higher-level structure as self-reports (e.g., McCrae & Mõttus, 2019). This will also help gauge the extent to which the hierarchical structure of VIA character strengths based on self-reports is biased by SDR. Although IPIP-VIA-R self-reports appear not to be biased by SDR in a detrimental way (see SOM on OSF), this point deserves greater attention in future research.

Furthermore, our analyses focused exclusively on the internal structure (i.e., constitution and content) of the higher-level VIA dimensions as well as their robustness and replicability across cultures. However, further thoughts and tests on the theoretical and practical utility of Level III, and also Level II, are needed. Therefore, future research should locate the dimensions of the Levels II and III in a nomological network of Dynamism/Plasticity and SSR/Stability, the Big Five and the Big Six/HEXACO, and agency and communion. Furthermore, it should demonstrate (incremental) predictive power of the Level-III-dimensions beyond these related constructs to gauge the added value of the three global VIA dimensions for applied purposes. Moreover, future research should show how the three global VIA dimensions can contribute to a better understanding of the sources of individual differences in character (e.g., by identifying neuronal and genetic correlates of the three dimensions) and their functionality/adaptivity (e.g., by theorizing why individual differences in the three dimensions were preserved in the evolutionary process, how individual differences in the three dimensions contribute to the functioning of societies and cultures). Also, to understand the role of the three global VIA dimensions in the course of people’s lives (marked by different stages of development and critical life events), it will be important to investigate their precursors and their potentially different developmental trajectories.

Finally, although the 24 VIA strengths represent an extensive trait space of well-defined and both theoretically and empirically supported constructs, they may not exhaust the entire character space that one could conceive: The VIA framework is *not* based on an exhaustive lexical approach. It restricts itself to cross-culturally, potentially even universally valued character traits thereby disregarding culture-specific character strengths that may matter in narrower contexts. Thus, our results must not be misunderstood as having identified an *exhaustive* character trait hierarchy. Rather more modestly, they are aimed at advancing the establishment of the factor-analytic *VIA* trait hierarchy. Beyond that, further mapping out the full character space and establishing an exhaustive character trait hierarchy is an important end—to which the results of our Bass-ackwards analyses may contribute: Components that splinter off on the levels below Level III, thereby missing the globality criterion, may point to underrepresented content of the character trait space (see Goldberg, 2006). For example, spirituality and self-regulation were no marker strengths (i.e., did not load highly) of any of the global dimensions on the Levels I–III across countries. Instead, they splintered off (e.g., as VI.4 and VI.5). There might be character strengths that correlate with spirituality or self-regulation and that are not fully represented yet in the (VIA) character trait space. Expanding the character trait space by such further character strengths, might potentially lead to the emergence of additional global dimensions (i.e., aggregates of the character strengths). Peterson and Seligman (2004) themselves expected that their collection of 24 character strengths might be extended in the future. Even so, it should be noted that these authors' theoretical groundwork was extensive and inclusive. As a result, we know of no study that identified relevant content that fit Peterson and Seligman's (2004) definition of character strengths but was absent from their classification more than 15 years since the initial publication of their classification. Thus, the initial list of 24 VIA character strengths has so far stood the test of time.

Conclusion

Our study suggests that a maximum of three interpretable and cross-culturally replicable global dimensions can be identified in the VIA trait space (as measured with VIA-IPIP-R). The three dimensions can be described as follows: Dimension III.1 (*positivity*) is characterized by forgiveness, zest, hope, and capacity for love, III.2 (*dependability*) is characterized by prudence, modesty/humility, integrity, and equity, and III.3 (*mastery*) is characterized by judgement, originality, perspective, valor, leadership, and social intelligence. Level III thus potentially represents the most useful level of aggregation on which to capture the essence or common core of the 24 VIA character strengths. This level of aggregation lends itself especially for empirical applications in which a parsimonious, global description of the VIA trait space or the prediction of broad outcomes is the goal. For an even more global description, Level II with its two meta-traits resembling the “Big Two” of Dynamism and Social Self-Regulation may also present a viable alternative. Future studies may develop scales to measure the two or three key dimensions in research contexts that do not require or allow the level of detail provided by the 24 original character strengths. Overall, our findings may contribute to the long-term goal of reaching consensus on the hierarchical structure of VIA and establishing a factor-analytic VIA—or even character—trait hierarchy, analogous to the personality trait hierarchy and the intelligence trait hierarchy.

Endnotes

[1] Throughout this paper, the term “level” refers to a hierarchical level in the trait space as identified through Goldberg’s (2006) Bass-ackwards analysis. The term “dimension” refers to a factor or principal component on that level. For example, Level III has three dimensions, whereas Level VI has six.

[2] Note that the term “hierarchy” has different meanings in different contexts. Peterson and Seligman (2004) use the term to describe their hierarchical theoretical classification of six abstract core virtues and 24 more specific character strengths. By contrast, in factor-analytic literature, the term is often used to describe the trait hierarchy of individual difference constructs such as personality or intelligence that results from the Bass-ackwards approach or similar techniques (e.g., Goldberg, 2006; Lang et al., 2016). Because the term “hierarchy” is completely agnostic as to how the hierarchy was constructed (i.e., through theoretical classification or through a factor-analytic approach) and equally appropriate in both contexts, we also use it in both contexts in the present paper. However, the study of the present paper is exclusively concerned with “solutions-hierarchies” created by the Bass-ackwards approach.

[3] Peterson and Seligman (2004) seem to assume that there are certain thresholds above which a person can be considered to possess a virtue: "We speculate that all these virtues must be present at above-threshold values for an individual to be deemed of good character. [...] We are comfortable saying that someone is of good character if he or she displays but 1 or 2 strengths within a virtue group" (p.13). In this context, Ruch and Proyer (2015, p. 3) state that "[a] core challenge is to define and validate a criterion for the presence of a strength".

[4] Of note, this statement only relates to tendencies implied by the factor-analytic model at the sample level. Not all individuals’ idiosyncratic response patterns will conform to the model. It is possible for individuals to score low on one character strength yet high on another, even though both strengths may load strongly on the same higher-level dimension.

[5] On a side note, we know of only two factor-analytic studies (Ng et al., 2017; Vanhove et al., 2016) that provided evidence in favor of a six-dimensional solution that matched Peterson and Seligman's (2004) theoretical classification by using confirmatory approaches or combining exploratory and confirmatory approaches. Ng et al. (2017) developed a bifactor model that was putatively able to recover the six virtues. As manifest variables, they used a 107-item subset of the VIA-IS, which they selected to maximize the unidimensionality of scales but without considering their content validity. Furthermore, they specified a global positivity factor, 24 character strengths factors, and represented the six core virtues as cross-loadings between items of character strength scales that were mapped onto the same virtue by Peterson and Seligman (2004). Vanhove et al. (2016) used an ultra-short 24-item measure of the VIA character strengths and achieved acceptable fit for a bifactor model with a global factor and six virtue factors only after removing three of the character strengths items. Thus, evidence for a six-dimensional solution recovering Peterson and Seligman's (2004) theoretically derived core virtues is limited and tenuous. As noted earlier, there is no inherent necessity for factor-analytic results to resemble Peterson and Seligman's theoretical classification.

[6] The implementation of the Bass-ackwards procedure by Waller directly computes the between- and within-level correlations of components from rotation matrices. Unlike the original Goldberg method, which computes correlations based on estimated component scores, the Waller method can also be applied when using factors from EFA or PAF, thus overcoming the indeterminacy problem of EFA-factor scores (Lang et al., 2016).

[7] With $\lambda = .50$ as lower bound for marker strengths we ensured that a component explained at least a quarter of the variance of its marker strengths. We chose $|\lambda| = .30$ as lower bound for co-defining strengths as this corresponds to a widely used rule of thumb in the interpretation of a factor/component.

[8] To ensure that the target rotation algorithm finds the global optimum, we manually rearranged the loading matrices before target rotation, so that corresponding components in Germany and UK (1) were aligned (i.e., not rotated to the opposite direction representing a virtue component in one and a vice component in the other country) and (2) represented the same column in both loading matrices. We also used a function that assures finding the global optimum with the help of random initial loading matrices. However, this function uses orthogonally rotated random initial loading matrices and did not operate reliably in our use case of obliquely rotated components. To our knowledge, no function using obliquely rotated random initial loading matrices is provided yet.

STUDY 3

Abstract

The Values in Action (VIA) framework encompasses 24 universally valued character strengths. Recent factor-analytic work has identified three global core strengths (metatraits) that proved to be well-interpretable and cross-culturally replicable: positivity, dependability, and mastery. So far, there are no scales to measure these core strengths. In the present study, we applied an Ant Colony Optimization algorithm to select three 6-item scales from the 96-item IPIP-VIA-R inventory. Thereby, we constructed balanced-keyed scales that cover the heterogeneous constructs well, showed good model fit and reliability across six samples from Germany and the UK (total *N* = 2,754), and achieved scalar measurement invariance across countries. Furthermore, we demonstrated the scales' validity by locating the three core strengths in a nomological net with personality and value metatraits, life satisfaction, and behavioral criteria. Available in the public domain, these both valid and economic core strength scales may further stimulate integrative research on personality and values.[§]

Keywords: VIA, character strengths, core strengths, metatraits, Ant Colony Optimization, economic scales, public domain instrument

[§] Study 3 was informally published as Partsch, M. V., Olaru, G., & Lechner, C. M. (2022). *Measuring global character dimensions: An ant colony optimization approach towards three core strengths scales*. PsyArXiv. https://doi.org/10.31234/osf.io/tdjh4

Introduction

When moral aspects of human behavior are of interest, researchers often focus on positive ethical traits such as character strengths and virtues (Fowers et al., 2021). Two decades ago, Peterson and Seligman (2004) introduced a comprehensive collection of 24 theoretically derived and empirically supported “values in action” (VIA): ubiquitously and morally valued, malleable character strengths. Locating them in a trait hierarchy, the 24 VIA character strengths are on a fine-grained level of abstraction similar to personality *facets* (McGrath et al., 2020; Partsch et al., 2022). That is, they are rather specific, narrow traits. However, we also require highly aggregated global VIA traits to understand how the VIA trait space is related to other trait spaces (e.g., the Big Five trait space), to parsimoniously describe human character, to predict global outcomes, and to allow for efficient character measurement.

In the lexical tradition of personality research, global personality traits are well-established, such as the Big Five and their metatraits, the Big Six, and the Big Two (e.g., DeYoung, 2006; McCrae & John, 1992; Saucier et al., 2014; Thalmayer & Saucier, 2014). Contrariwise, there was no consensus regarding factor-analytically derived global VIA (meta-)traits[1] for a long time (see review in Partsch et al., 2022). Several recent studies, however, converge on the notion that three global metatraits can be identified in the VIA trait space. Most recently, a comprehensive factor-analytic study into the hierarchy of the VIA trait space by Partsch et al. (2022) based on the public domain International Personality Item Pool (IPIP; Goldberg et al., 2006) identified three well-interpretable, cross-culturally replicable, global core strengths: positivity, dependability, and mastery. However, there are no scales to measure these core strengths so far.

In the present study, we developed three 6-item scales for the measurement of these newly identified core strengths in a total of six samples from Germany and the UK (total *N* =

2,754) based on the 96 public domain items from the IPIP-VIA-R inventory (Bluemke et al., 2021). We used the algorithmic Ant Colony Optimization (ACO) approach (e.g., Olaru et al., 2019; Schroeders et al., 2016) to optimize the item selection in terms of balanced keying, measurement model fit, factor saturation, and scalar measurement invariance. Furthermore, we evaluated the nomological network of the three core strengths and examined their criterion validity.

VIA—Theoretical Classification and Factor-Analytic Trait Hierarchy

24 VIA Character Strengths and Their Theoretical Classification

The building blocks of the VIA trait space are 24 character strengths, which are theoretically underpinned and empirically supported individual difference constructs. Peterson and Seligman (2004) presented these 24 character strengths as part of a classification of "good character" in their "manual of the sanities", which they introduced as positive-psychological counterpart to classification manuals of mental diseases and disorders. The 24 character strengths resulted from a comprehensive collection of positive characteristics that Peterson and Seligman (2004) subsequently winnowed based on different inclusion criteria (e.g., traits that are cross-culturally and morally valued, fulfilling for the self and others, and socially/institutionally cultivated).

Peterson and Seligman (2004) classified the character strengths by mapping them onto six more abstract, universally acknowledged core virtues (wisdom, courage, humanity, justice, temperance, and transcendence). Importantly, their Linnaean classification resulted from a purely theoretical approach: The authors (1) identified the six virtues based on an extensive review and theoretical analysis of philosophical and religious writings and (2) based the strength-virtue-mapping on a theoretical analysis of common attributes of the strengths (e.g., they mapped "cognitive strengths" such as creativity and curiosity onto the virtue of wisdom).

Three Core Strengths in the Factor-Analytic Trait Hierarchy

A different approach to map specific traits onto global traits is to establish a factor-analytic trait hierarchy. Thus, researchers determine empirically by means of factor-analytic methods such as Goldberg's (2006) Bass-ackwards analysis the number and nature of global constructs that reside on more abstract levels in the trait hierarchy. These factor-analytically derived global constructs summarize the essence and the common core of the underlying more specific constructs—as expressed in their (co-)variation. This is by far the most common approach in individual difference research. For example, in the Big Five personality trait space, facets (e.g., Soto & John, 2017a) reside on a similarly low abstraction level like the 24 character strengths in the VIA trait space. These facets are summarized in the Big Five factors of personality which again can be summarized in two metatraits (i.e., Plasticity and Stability; DeYoung, 2006).

For the VIA trait space, such global dimensions or metatraits have not been successfully established until recently. Various factor-analytic approaches to aggregate the 24 character strengths were of varying methodological quality and yielded inconsistent results that ranged from anywhere between 1 and 6 global dimensions (for a detailed review see Partsch et al., 2022). However, several more recent studies clearly pointed to three global dimensions (especially McGrath, 2015; McGrath et al., 2018, 2022; McGrath & Wallace, 2021). McGrath and colleagues interpreted the three factors they identified as *Caring* (supposed to capture interpersonal strengths/interpersonal concern), *Inquisitiveness* (intellectual strengths/information seeking), and *Self-Control* (intrapersonal strengths/intrapersonal regulation) and referred to them as "three-virtue model".

To further contribute to the establishment of the factor-analytic trait hierarchy, Partsch et al. (2022) revisited the hierarchical structure of the 24 character strengths with a refined methodological approach. They scrutinized Levels I–VIII of the trait hierarchy that resulted from applying Goldberg's Bass-ackwards analysis to data from large heterogeneous samples of respondents from Germany and the UK who completed a revised VIA inventory based on the IPIP (IPIP-VIA-R; Bluemke et al., 2021). They found that only Levels I–III consisted exclusively of well-interpretable, truly global (i.e., comprising at least three highly loading character strengths), and cross-culturally replicable dimensions, whereby Level III emerged as the most useful level of analysis. This level comprises the three global VIA dimensions or metatraits *positivity*, *dependability*, and *mastery*. Because these dimensions each summarize the common core of different character strengths, we will refer to them as the *three core strengths* in the remainder of this article. In Table 3.1, we illustrate the three core strengths and present both their marker strengths and related global constructs.

Despite some overlap, the three core strengths by Partsch et al. (2022) and the three-virtue model (e.g., McGrath et al., 2018) represent different constructs bearing different labels. These differences result from different inputs and methodological choices in the factor analyses (PCA using the IPIP-VIA-R in Partsch et al., 2022, versus, for example, PAF using the VIA-IS-R in McGrath & Wallace, 2019 and McGrath et al., 2022) and different interpretive rationales (abstracting the common core of the underlying character strengths following the factor-analytical logic in Partsch et al., 2022, versus attempting to bring the three factors in line with common virtue concepts and other virtue theories as explained in McGrath et al., 2018).

Table 3.1

The Three Core Strength Constructs

	Positivity	**Dependability**	**Mastery**
Marker strengths[a]	**forgiveness**, **zest**, **hope**, and **capacity for love**	**prudence**, **modesty/humility**, **integrity**, and **equity**	**judgment**, **originality**, **perspective**, **valor**, **leadership**, and **social intelligence**
Illustration	A person scoring high on positivity can be characterized as a **cheerful, optimistic, and forbearing conciliator**: They believe in the good in people and have affectionate, trusting relationships, are full of zest for life, and look on the bright side of things.	A person scoring high on dependability can be characterized as a **reliable, trustworthy, and caring "tower of strength"**: They have a realistic and accurate view on themselves, others, and situations or events. They translate their accurate evaluations into careful decisions and trustworthy, fair, and considerate interpersonal dealings.	A person scoring high on mastery can be characterized as a **wise and ingenious leader**: They have a sharp and flexible mind that translates into confident behavior and a commanding manner. They master (new) information and interpersonal dealings.
Related constructs	Positivity[b], Stability[d,f]	Social Self-Regulation[c], *Stability*[d,f], Communion[e]	Dynamism[c], Plasticity[d], Agency[e]

Note. Partly taken from Partsch et al. (2022). [a]The marker strengths of a core strength are the VIA character strengths that loaded $\lambda \geq .50$ exclusively on the respective core strength in both Germany and the UK in the Bass-ackwards analyses by Partsch et al. (2022). [b]Caprara et al. (2012), [c]Thalmayer and Saucier (2014), [d]DeYoung (2006), [e]Bakan (1966). [f]Partsch et al. (2022) assumed the strongest relationship between Stability and dependability. In the present study (see Table 3.6), we found an even stronger relationship between Stability and positivity.

Measurement of the Three Core Strengths

How Many Items are Feasible?

Soto and John (2019) identified a scale length of 6–9 items as the "sweet spot" for the measurement of global traits. With 6–9 items, one can usually ensure adequate construct coverage, reliability, and validity while maintaining the scales' efficiency and essential unidimensionality and minimizing idiosyncratic, systematic response biases (e.g., McCrae &

Mõttus, 2019; Soto & John, 2019). There are several examples of scales that measure global constructs or metatraits with 6–9 items: Soto and John (2017b) developed a 6-item scale for each Big Five domain in the BFI-2-S, Thalmayer and Saucier (2014) measured Dynamism and Social Self-Regulation (i.e., the Big Two metatraits from the Big Six framework) with seven items each, and Entringer et al. (2021) used 6–7 items to measure agency and communion. To assess the three core strengths adequately, we opted for a scale length of six items per core strength. This would allow the three scales to cover each of the 4–6 marker strengths per core strength (displayed in Table 3.1).

How to Select Items for Scales of Global Traits?

Ensuring Good Construct Coverage. Usually, items for measuring global traits are not newly generated but selected from larger inventories that were originally constructed to measure the more specific traits (e.g., facets) underlying these global traits. Items from each facet are included to ensure good construct coverage. For example, Soto and John (2017b) constructed BFI-2-S's 6-item scales per Big Five domain by selecting two out of four items of each facet scale of the BFI-2 (Soto & John, 2017a). Likewise, Thalmayer and Saucier (2014) selected 1–3 items from each Big Six dimension to construct two scales measuring the Big Two metatraits. In these examples, each item could only be selected for a single global trait scale (e.g., Sociability, Assertiveness, and Energy Level items could only be selected for Extraversion but for none of the other Big Five domains).

The situation is somewhat more complex for the three core strengths. As per Partsch et al.’s (2022) results, core strengths are each characterized by their cross-national, unique set of highly loading *marker strengths* that have loadings of $\lambda \geq .50$ exclusively on the respective dimension in both Germany and the UK (see Table 3.1). Thus, the three core strengths mainly represent the essence of 14 of the 24 VIA character strengths. The remaining 10 *co-defining* strengths (appreciation of beauty, citizenship/teamwork, curiosity, gratitude, humor, industriousness/perseverance, kindness, love of learning, self-regulation, and spirituality/religiousness) loaded $|\lambda| \geq .30$ on at least one of the three core strengths in both countries (but never $\lambda \geq .50$ on the single same dimension across Germany and the UK, which defines the marker strengths).[2] From this Partsch et al. (2022) concluded that "none of the character strengths fell outside the trait space covered by the three dimensions" (p. 13). From this we concluded that the co-defining strengths should not be completely neglected in the development of the three core strength scales, but the marker strengths should receive greater weight. We thus derived the following construction rationale: In a 6-item core strength scale, each of its 4–6 marker strengths is represented with at least one item. This ensures that the scales indeed capture the abstract constructs of positivity, dependability, and mastery. In addition, single items from co-defining strengths might lend themselves to complement the operationalization of these abstract constructs. In Figure S1, that we provide as Supplementary Online Material (SOM) on OSF (https://osf.io/pdntc/), we visualize how *core* strengths, *character* strengths, *marker* strengths, and *co-defining* strengths are related.

ACO—Simultaneously Optimizing Multiple Psychometric Properties. In addition to construct coverage or content validity, scales must satisfy psychometric criteria such as reliability, model fit, and measurement invariance across countries or languages. Identifying which combination of items best fulfil these multiple psychometric criteria is a complex combinatorial problem. Solving it by estimating all possible 18-item models is computationally too demanding. We, therefore, used the heuristic item selection algorithm ACO to approach the optimal solution (Olaru et al., 2019; Schroeders et al., 2016). ACO selects and evaluates item combinations across several iterations until it finds an optimal or close-to-optimal solution. It starts by selecting random item sets from the item pool and evaluates each set based on multiple user-defined psychometric criteria. The items of the item set that meets the criteria best receive an increase in selection probability for subsequent iterations. We let ACO repeat this process until it yields no further improvement in the psychometric criteria.

Previous studies (e.g., Jankowsky et al., 2020; Olaru & Danner, 2021; Olaru & Jankowsky, 2021) used ACO for short scale construction (i.e., selecting items from a long version of a scale to construct its short version). In short scale development each long scale represents the clearly defined item pool for the selection of the short scale. Contrariwise, the present item selection scenario (selecting items for VIA metatraits) is more complex because the items from the co-defining strengths cannot be uniquely assigned for the selection of a single core strength scale. We therefore expanded the ACO algorithm by a new item sampling feature that allows ACO to choose items from co-defining strengths for different scales. The higher complexity of the present selection scenario also involves some human judgement of the ACO-based results (more details below). Thus, we slightly adjusted the classic ACO approach to short scale development to tailor it to the requirements of developing scales for (VIA) metatraits.

The Present Study

To summarize, we developed and validated an 18-item inventory to measure three global core strengths (or metatraits) based on the public domain IPIP-VIA-R inventory (Bluemke et al., 2021), which is based on the IPIP (Goldberg et al., 2006). We approached the combinatorial problem of selecting the most suitable 18 out of 96 items with an ACO algorithm (Olaru et al., 2019; Schroeders et al., 2016). To assess the robustness of our results, we used six heterogeneous samples from three data collections each conducted in Germany and the UK, allowing for extensive cross-validation and built-in replication.

We developed the scales in three steps. In the first step, we fed data from the first data collection (samples 1 and 2, hereinafter referred to as *training samples*) into the ACO algorithm to select a set of 18 items (i.e., six per scale) that would approach maximum construct representation, balanced keying, measurement model fit, factor saturation, and measurement invariance across Germany and the UK. In the second step, we re-examined the psychometric properties of the ACO item selection based on data from the second data collection (samples 3 and 4, hereinafter referred to as *validation samples*) to check if the item selection optimized based on the training samples can be generalized and does not represent a sample-specific, thus overfitted selection (Olaru et al., 2019). In the third step, we used data from our third data collection (samples 5 and 6, hereinafter referred to as *use-case samples*) to examine whether the scales maintain their psychometric properties when administered as an independent 18-item inventory instead of being embedded in the original item pool of 96 items (as recommended by Smith et al., 2000).

To validate the newly developed scales, we located them in the nomological net of the Big Five metatraits Plasticity and Stability (DeYoung, 2006; DeYoung et al., 2002) and the metatraits abstracted from Schwartz' (1992, 1994) Theory of Basic Human Values (i.e., social growth orientation and personal growth orientation; Rudnev et al., 2016; Schwartz, 1992, 2012). Furthermore, we looked at the association of the three core strengths with several criteria: general life satisfaction and various behavioral correlates.

Method

Open Data & Material

We provide the complete data and R code including a renv-lockfile for reproducibility (Ushey, 2020) on the project website on OSF. We conducted all analyses in R (version 4.0.2; R Core Team, 2020).

Data

We used data from six samples from Germany (samples 1, 3, and 5) and the UK (samples 2, 4, and 6). We collected the *training samples* (i.e., samples 1 & 2) in 2018, the *validation samples* (i.e., samples 3 & 4) in 2019, and the *use-case samples* (i.e., samples 5 &

6) in 2021. For each data collection, we contracted the commercial online access panel provider respondi AG. For Germany, we drew quota samples based on gender, age, and education (obtained from Census 2011 for the surveys of 2018 and 2019 and from Census 2017 for the survey of 2021, respectively). For the UK, we drew parallel samples. In each data collection, we invited a subsample of the main survey to a retest survey 2–3 weeks after completion of the main survey. To ensure data quality, we excluded careless responders separately per sample 1–6.[3] In Table 3.2, we report both the characteristics of all analysis samples 1–6 (i.e., after case exclusion) and the sizes of all six samples before and after case exclusion and in the retest. Furthermore, we provide an overview of information on and use of the six samples in Table S1 as SOM on OSF. With $N \geq 420$, we deemed our analysis samples of suitable size for all our analyses (e.g., prevention of overfitting in ACO item selection, adequate for measurement invariance testing; Chen, 2007; Olaru et al., 2019).

Missing data were negligible in the first and second data collection (i.e., samples 1–4) ranging between 0.00–0.60% across all items from VIA and validation scales. In the third data collection (i.e., samples 5–6), we applied a 3-form planned missingness design (3F-PMD; J. W. Graham et al., 1996, 2006) to reduce response burden given the large number of administered scales. Thus, we prepared three questionnaire forms each containing only two thirds of the items of several instruments and randomly assigned one of these forms to each respondent. In 3F-PMD, the items are assigned to questionnaire forms in such a way that half of the presented items overlap with one of the other forms, while the other half overlaps with the other. From the newly developed core strength scales, we presented two of three scales to each form. Introducing missings completely at random (MCAR) and in a way that missing information can be drawn from the other two forms, the 3F-PMD allows conducting analyses with full information maximum likelihood (FIML). Except for the estimation of the test–retest correlations (r_{tt}), we conducted all analyses using FIML.

Table 3.2

Sample Characteristics of Samples 1–6

	Training samples (1st data collection)		Validation samples (2nd data collection)		Use-case samples (3rd data collection)	
	Germany (Sample 1)	UK (Sample 2)	Germany (Sample 3)	UK (Sample 4)	Germany (Sample 5)	UK (Sample 6)
N						
Before case exclusion	518	522	509	524	463	483
Analysis samples	476	474	468	476	420	440
Retest samples	117	107	111	118	202	199
Age in years						
M	44.23	44.45	43.35	44.15	43.63	44.43
(*SD*)	(15.08)	(14.31)	(14.76)	(14.39)	(13.98)	(13.11)
[*range*]	[18–69]	[18–69]	[18–69]	[18–69]	[18–65]	[19–65]
Proportion of women (%)	51.47	50.42	50.21	52.31	51.19	49.77
Educational level (%)						
Low	34.66	33.54	34.19	33.61	25.95	24.32
Intermediate	32.98	34.39	32.69	32.98	33.81	32.95
High	32.35	32.07	33.12	33.40	40.24	42.73

Note. Educational levels: low: no educational qualification, lower secondary leaving certificate; intermediate: intermediate school leaving certificate; high: higher education entrance qualification.

Measures

VIA Item Pool. We developed the three core strength scales based on the IPIP-VIA-R inventory. The 96 items of the IPIP-VIA-R (Bluemke et al., 2021) measure each of the 24 VIA character strengths (Peterson & Seligman, 2004) with two positively and two negatively keyed items on a fully labeled 5-point rating scale ranging from “does not apply at all” to “applies completely”.

Personality Metatraits. We applied measures of the metatraits from the Big Five framework: *Stability* and *Plasticity* (DeYoung, 2006; DeYoung et al., 2002). Whereas Stability expresses the common core of Agreeableness (i.e., stable social relationships), Conscientiousness (i.e., stable goal pursuit), and Emotional Stability (i.e., absence of negative affect), Plasticity expresses the common core of Extraversion and Open-Mindedness—namely, a flexible, explorative, and positive affective approach to the world. While Stability facilitates social integration, Plasticity facilitates personal growth. We measured Stability and Plasticity with the BFI-2-XS (Rammstedt et al., 2020; Soto & John, 2017b). The 15 items of the BFI-2-XS measure each Big Five dimension with three items (i.e., one from each facet), of which one or two are negatively keyed, on a 5-point rating scale ranging from "strongly disagree" to "strongly agree". We administered the BFI-2-XS in the validation samples. We computed Stability scores as unit-weighted means of the Agreeableness, Conscientiousness, and Emotional Stability (i.e., recoded Negative Emotionality) items and Plasticity scores as unit-weighted means of the Extraversion and the Open-Mindedness items. The internal consistencies of the scores were Cronbach's $\alpha_{Germany} = .70/\alpha_{UK} = .73$ for Stability and $\alpha_{Germany} = .59/\alpha_{UK} = .58$ for Plasticity.

Value Metatraits. We applied measures of the two higher-order value dimensions from the basic human values framework: On the one hand, the preference for self-transcendence values over self-enhancement values, that we labeled s*ocial growth orientation* based on the representations in Schwartz (1992, 2012) and Rudnev et al. (2016); on the other hand, a preference for openness to change over conservation, which we labeled *personal growth orientation* accordingly. Social growth orientation and personal growth orientation both express the preference for anxiety-free growth values over anxiety-based self-protection values. Social growth orientation represents an approach- and growth-oriented regulation of one's interpersonal relationships, in which one transcends selfish interests and enhances the others. Personal growth orientation represents an approach- and growth-oriented regulation of one's personal interests and characteristics, which involves striving for a thrill of pleasure, novelty, and mastery (Schwartz, 2012). We measured social growth orientation and personal growth orientation with the 21-item version of the Portrait Values Questionnaire (PVQ-21; Schmidt et al., 2007; Schwartz, 2003a). The PVQ-21 measures each of the ten basic human values with 2–3 items on a 6-point rating scale ranging from "very much like me" to "not like me at all" (including the response option "don't know"). We administered the PVQ-21 in the validation samples. Following the approach described on ESS EduNet (n.d.), we subtracted the self-enhancement scores (i.e., unit-weighted mean score of power and achievement items with $\alpha_{\text{Germany}} = .77$ and $\alpha_{\text{UK}} = .79$) from the self-transcendence scores (i.e., unit-weighted mean score of universalism and benevolence items with $\alpha_{\text{Germany}} = .81$ and $\alpha_{\text{UK}} = .79$) to compute social growth orientation scores. To compute personal growth orientation scores, we subtracted the conservation scores (i.e., unit-weighted mean score of conformity, tradition, and security items with $\alpha_{\text{Germany}} = .68$ and $\alpha_{\text{UK}} = .74$) from the openness to change scores (i.e., unit-weighted mean score of hedonism, stimulation, and self-direction items with $\alpha_{\text{Germany}} = .77$ and $\alpha_{\text{UK}} = .75$).

General Life Satisfaction. We used a measure of general life satisfaction to assess respondents' global evaluation of their subjective well-being (Diener, 1984).We applied the single-item L-1 scale (Beierlein et al., 2015; Nießen et al., 2020) on an 11-point rating scale ranging from "not at all satisfied" to "completely satisfied". We administered the L-1 in the use-case samples.

Behavioral Correlates. We used 22 items from a newly developed battery of behavioral correlates that allows to measure the frequency of a broad range of behaviors. We bundled the single items into seven thematic behavioral clusters (see Table 3.6 and OSF project website). The behavioral items are inspired by the behaviors presented in Schwartz

(2015; Table 3). The different behaviors can be connected to different underlying basic human values. Accordingly, based on Schwartz’ (2015) empirical results or theoretical reasoning, we allocated the seven thematic behavioral clusters in Schwartz’ value circle (e.g., Schwartz, 1994): (1) self-enhancing behaviors (primarily motivated by the basic human value of power), (2) risk-taking behaviors[4] (hedonism and stimulation), (3) personal growth behaviors (self-direction), (4) activism and civil courage behaviors (self-direction and universalism), (5) sustainable behaviors (universalism), (6) charitable behaviors (universalism and benevolence), and (7) provision-making behaviors (security). We presented the items to the use-case samples with a 5-point scale ranging from “never” to “very often” (including the response option “prefer not to say”).

Statistical Analyses

Scale Development

In the first step, we used ACO to select 18 (i.e., six items per core strength) out of 96 IPIP-VIA-R items based on the training samples. In a nutshell, ACO selects and evaluates several item combinations (in this case 18-item models) based on user-given optimization criteria (e.g., model fit). Over the course of several iterations, ACO learns which items are best suited to maximize the given optimization criteria and will return a (close-to-)optimal solution without requiring an estimation of all possible item combinations (for detailed information, see Olaru et al., 2019; Schroeders et al., 2016). In the following, we describe the properties of the item selection used in this study, namely the (a) the measurement model, (b) the item sampling procedure, (c) optimization criteria, and (d) the ACO parameter settings.

(a) Measurement Model. As measurement model, we specified a correlated first-order factor model (Brunner et al., 2012) with three factors representing the three core strengths. In addition, we included a method factor capturing acquiescent response style (ARS), also known as “yeah-saying”, that is, the tendency to agree to survey items independent of their content and keying (Jackson & Messick, 1958). ARS is a wide-spread source of bias in survey data especially from cross-national surveys (Lechner et al., 2019; Rammstedt et al., 2013) and can be corrected for in confirmatory factor analysis (CFA) via a latent ARS factor (Billiet & McClendon, 2000). All items loaded on the ARS factor with these loadings fixed to 1 (for both positively and non-recoded negatively keyed items) and the factor variance freely estimated. Figure 3.1 illustrates the measurement model. We used multi-group CFA to test measurement invariance across countries (i.e., Germany and the UK). More specifically, we compared the configural (i.e., no parameter constraints across groups), the metric (i.e., equal factor loadings across groups) and the scalar measurement invariance level (i.e., equal factor loadings and item intercepts across groups). We estimated all models with robust maximum likelihood (MLR) estimation using the R-package lavaan (version 0.6-7; Rosseel, 2012).

(b) Item Sampling. In previous studies that used ACO for item selection, all items from the item pool were uniquely assigned to one factor for which ACO could select them (e.g., Olaru & Danner, 2021; Olaru & Jankowsky, 2021). In the present study, we applied this logic to the 56 marker strengths items. That is, we preset that ACO could select items measuring forgiveness, zest, hope, and capacity for love (i.e., the marker strengths of positivity) only for the positivity factor. Likewise, we preset that ACO could select items measuring prudence, modesty/humility, integrity, and equity (i.e., the marker strengths of dependability) only for the dependability factor and items measuring judgment, originality, perspective, valor, leadership, and social intelligence (i.e., the marker strengths of mastery) only for the mastery factor.

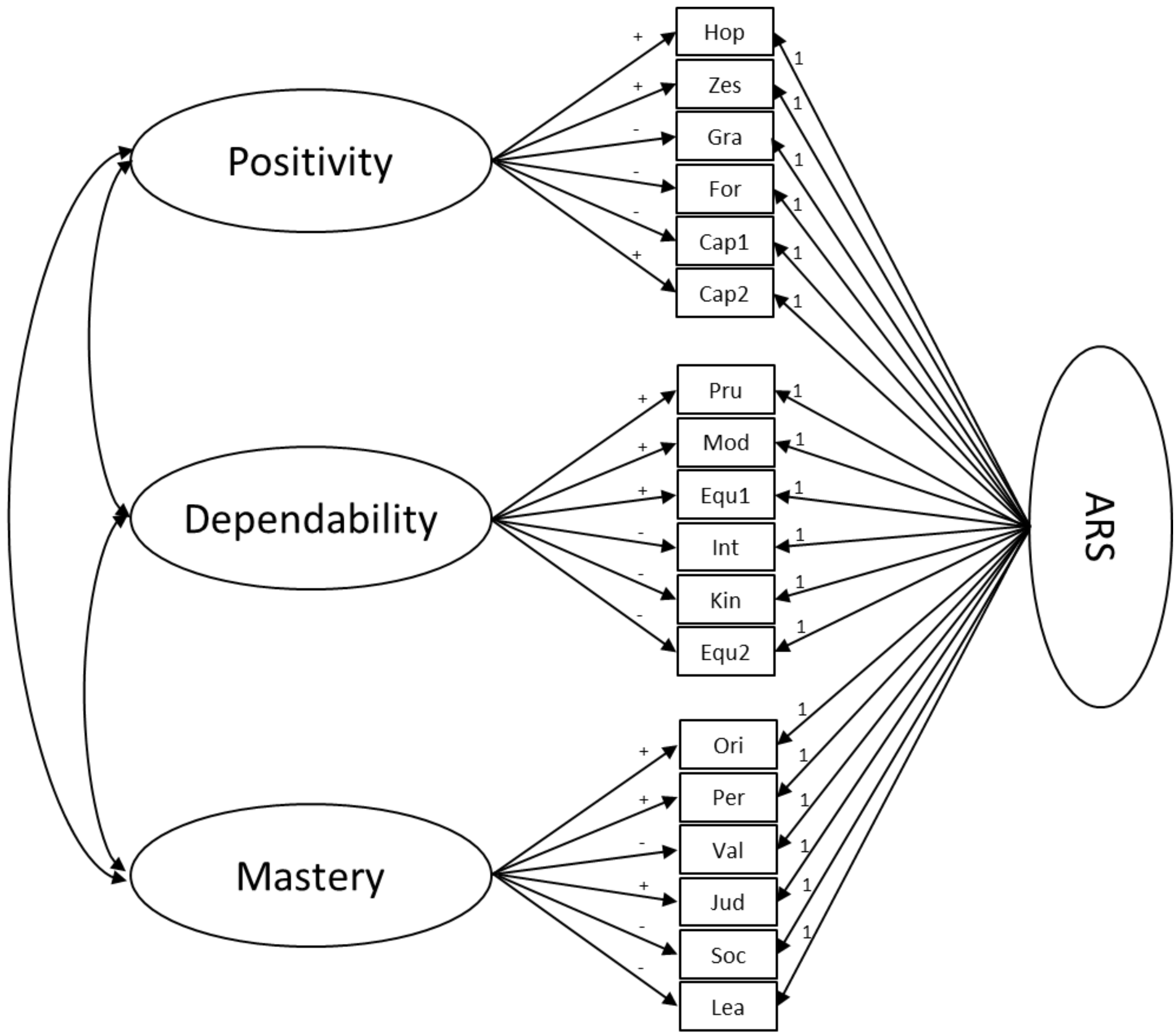


Figure 3.1. Measurement model of the three core strength scales. ARS = acquiescent response style. Items corresponding to the item labels in the boxes can be found in Table 3.3. +/− signs indicate item keying and positive/negative factor loading, respectively. Loadings of acquiescence factor were fixed to 1.

Unlike in previous studies using ACO for item selection, the IPIP-VIA-R item pool also comprises 40 items measuring 10 co-defining strengths, that, by definition, cannot be uniquely assigned to be selected for a single core strength. At the same time, these items cannot be fully neglected in the measurement of the three core strengths either. To do justice to this, we extended the previous ACO sampling procedure to also include items from co-defining strengths. We programmed the item sampling as follows:

For each item selection, ACO had to draw at least one item of each marker strength. That is, four out of six items of the positivity and dependability scale, respectively, and (all) six items of the mastery scale, were marker items. In this way, we made sure that each of the three scales would measure its core strength's definitional core. Selecting six items per scale allowed to choose two more items for both the positivity and the dependability scale. Here, ACO was free to select items from either the core strength's marker strengths or from the co-defining strengths, while optimizing the criteria outlined below. That is, we let ACO decide empirically whether any of the co-defining items lend themselves as indicators of positivity or dependability—and if so, which ones did.[5] ACO was not allowed to choose items from the same co-defining strength for different factors.[6]

(c) Optimization Criteria. Within the set measurement model and sampling procedure, ACO approached the optimal item selection by maximizing the criteria described below. We scaled all criteria to a range from 0 (minimum) to 1 (maximum) to ensure that ACO weighted them equally. For that purpose, we transformed the criterion for balanced keying with an exponential function and all other criteria with a logistic function with turning point at the critical cutoff (see R script on OSF for exact transformations; see Olaru et al., 2019 for an illustration of the criterion transformation). We computed the unit-weighted mean across the transformed criteria to obtain the overall optimization index.

Balanced Keying. This optimization criterion attained its minimum value of 0 when the item selection resulted in three fully unbalanced scales (i.e., each scale consisted of six either positively or negatively keyed items), and its maximum value of 1 if three fully balanced-keyed scales resulted (i.e., if each scale consisted of three positively and three negatively keyed items).

Model Fit. We optimized the fit of the scalar-invariant measurement model based on the comparative fit index (CFI) and the root mean squared error of approximation (RMSEA). Because good model fit does not require maximum values of CFI or RMSEA (i.e., CFI = 1.000 and RMSEA = 0.000), ACO aimed at reaching CFI ≥ .900 (as we expected medium-size factor loadings only that are generally associated with a smaller CFI; Moshagen & Auerswald, 2018) and RMSEA ≤ .060 (as recommended by Hu & Bentler, 1999).

Factor Loadings. ACO simultaneously aimed at reaching standardized factor loadings of $|\lambda| \geq .30$ (commonly used rule of thumb) and average variance extracted (AVE, i.e., average squared standardized factor loadings) of AVE ≥ .50 of each factor (as suggested by Fornell & Larcker, 1981).[7]

Measurement Invariance. To allow for cross-national comparisons, metric invariance is required; for mean-level comparisons, scalar invariance is required. We therefore aimed to obtain scales that reach scalar invariance across Germany and the UK. ACO minimized the decrease in model fit between the metric-invariant and the scalar-invariant model. Thereby, it aimed at reaching $\Delta CFI_{(metric - scalar)} < .010$ (as recommended by Chen, 2007).

(d) ACO Parameter Settings. Within each iteration, ACO compared 60 models, identified the best solution based on the overall optimization index, and increased the selection probability of the 18 items contained in the best solution by the overall optimization index. After each iteration, the selection probability of each item was reduced by 1% to decrease the probability that ill-suited items, that had been selected in the first iterations (i.e., the early learning phase of ACO), were re-selected in later iterations (“pheromone evaporation”).[8] ACO aborted the search after 40 iterations that did not yield a new best solution. Because ACO is a probabilistic item selection approach that might not always find the best possible solution, we ran ACO ten times with different random number generator seeds. Olaru et al. (2019) and Schroeders et al. (2016) provide detailed descriptions on how the ACO algorithm learns and proceeds.

Choice and Evaluation of Best Item Selection. Studies using ACO for item selection—to develop short scales from existent long scales—usually choose the solution of the ACO run that yielded the highest overall optimization index (e.g., Olaru & Danner, 2021;

Olaru & Jankowsky, 2021). However, because of the large number of criteria to be considered and the unrestricted item sampling from co-defining strengths, our item selection scenario—to develop scales of newly identified global constructs—was too complex and had too many degrees of freedom to fully computerize it. Therefore, we looked at each best solution of all ten ACO runs (we present them all as SOM on OSF) and evaluated them based on the following criteria: (1) The final solution should contain items from co-defining strengths that represent the core strength, for which ACO selected them, well. By definition, co-defining strengths could not be a priori (i.e., on *theoretical* grounds) assigned for the selection for a single factor. Hence, ACO decided *empirically* (i.e., with the aim to maximize the overall optimization index) if and for which scale it selected items from co-defining strengths. We then evaluated how each of the best solutions performed in terms of construct representation. (2) The final solution must be fully balanced-keyed. (3) It must show good model fit (i.e., CFI ≥ .900, RMSEA ≤ .060; furthermore, SRMR ≤ .080 as recommended by Hu & Bentler, 1999). (4) It must achieve scalar measurement invariance across Germany and the UK. As recommended by Chen (2007), we would reject the scalar measurement invariance model, if $\Delta CFI_{(metric – configural)} \geq -.010$ in combination with either $\Delta RMSEA_{(metric – configural)} \geq .015$ or $\Delta SRMR_{(metric – configural)} \geq .030$, and if $\Delta CFI_{(scalar – metric)} \geq -.010$ in combination with either $\Delta RMSEA_{(scalar – metric)} \geq .015$ or $\Delta SRMR_{(scalar – metric)} \geq .010$, respectively. Furthermore, decreases in BIC (i.e., an increasing balance between model fit and parsimony) from the configural to the metric and from the metric to the scalar model larger than 10 provide very strong evidence to accept the scalar invariant model (based on Raftery, 1995). (5) Factor correlations of the final solution must not be ≥ .80 in either country to make sure that the three core strengths indeed represented distinct latent constructs. Furthermore, in the final solution, (6) the minimum standardized loading on each factor in both countries should not fall below .30 and (7) McDonald’s ω (McDonald, 1999) of each scale should not fall below .60 in both countries.[9] We chose these lower boundaries for factor loadings and ωs because we expected only medium-size factor loadings (and some item uniqueness in return) and internal consistencies for our heterogenous 6-item scales.

Based on these criteria, we chose the solution with the 5th best overall optimization index as final core strength scales. This solution considerably overlapped with those from the four ACO runs yielding slightly higher overall optimization indices but outperformed them in terms of construct representation and (low) factor correlations.[10] In the second and third step, we evaluated the final item selection based on the validation samples and the use-case samples. Therefore, we applied the same evaluation criteria as for the training samples.

Scale Validation

To validate the newly developed core strength scales, we placed them in a nomological net with personality and value metatraits and relevant criteria. First, we looked at their associations with Stability and Plasticity—the two metatraits from the Big Five framework (DeYoung, 2006; DeYoung et al., 2002). Because readers may be more familiar with the Big Five dimensions than with their metatraits, we additionally provide the nomological net of the three core strengths and the Big Five as SOM on OSF. Second, we looked at the associations between the VIA scales and social growth orientation and personal growth orientation—the two metatraits from the basic human values framework (Rudnev et al., 2016; Schwartz, 1992, 2012). Third, we looked at the associations between the VIA scales and several criteria: general life satisfaction and different behavioral clusters that are based on Schwartz' basic human values (e.g., Schwartz, 1994).

For each validation construct or criterion, we hypothesized the strongest and, where feasible, second strongest correlation(s) among the three core strength scales. Readers can retrieve the pre-registration of our validation hypotheses from the OSF project website. To derive these hypotheses, we considered the extent of construct overlap between the validation constructs and the three core strengths, that is, we theorized to what extent they represent the same constructs fulfilling the same function (e.g., providing stability or facilitating flexibility) or having the same motivational basis (e.g., promoting growth of others or the self). Furthermore, we considered the extent to which the three core strengths and the validation criteria (i.e., life satisfaction, different behaviors) co-occur.

We used multi-group structural equation modeling (SEM) and estimated the models with MLR. In each SEM model, we included the scalar invariant measurement model of the three core strength scales and the manifest scale score of one or several validation constructs or criteria. To test our hypotheses, we looked at the correlations between the variables of interest in the SEM models. In the first model, we tested our validation hypotheses for Stability, Plasticity, social growth orientation, and personal growth orientation based on the validation samples. In the second model, we tested our hypotheses for general life satisfaction based on the use-case samples. Furthermore, we tested our hypotheses for the behavioral correlates based on both seven behavioral cluster scores and always the single behaviors these clusters consisted of in separate models (i.e., 14 models in total) based on the use-case samples. To handle the missing data from the 3F-PMD in the use-case samples, we included the seven behavioral clusters with pseudo-indicator models as manifest model-based composite scores in the SEM model (Rose et al., 2019).

Table 3.3

Items of the Three Core Strength Scales

IL	K	Items German	Items English
		Positivity	
Hop	+	Selbst bei Herausforderungen bleibe ich hoffnungsvoll.	Remain hopeful despite challenges.
Zes	+	Ich freue mich auf jeden neuen Tag.	Look forward to each new day.
Gra (CS)	−	Ich finde in meinem Leben nur wenige Dinge, für die ich dankbar sein könnte.	Find few things in my life to be grateful for.
For	−	Ich kann anderen nicht so leicht vergeben.	Find it hard to forgive others.
Cap1	−	Es fällt mir grundsätzlich schwer, Liebe anzunehmen.	Have difficulty accepting love from anyone.
Cap2	+	Ich bin mir gewiss, dass es Menschen in meinem Leben gibt, denen mein Wohl genauso wichtig ist wie ihr eigenes.	Know that there are people in my life who care as much for me as for themselves.
		Dependability	
Pru	+	Bei Entscheidungen gehe ich gerne auf Nummer sicher.	Make careful choices.
Mod	+	Niemand würde mich als arrogant bezeichnen.	Would never be described as arrogant.
Equ1	+	Mir ist es wichtig, dass alle Menschen die gleichen Rechte haben.	Believe that everyone's rights are equally important.
Int	−	Wenn ich mir dadurch Ärger ersparen kann, nehme ich es mit der Wahrheit nicht so genau.	Lie to get myself out of trouble.
Kin (CS)	−	Ich werde schnell ungeduldig, wenn andere mir von ihren Problemen erzählen.	Get impatient when others talk to me about their problems.
Equ2	−	Ich tue immer wieder mal Dinge auf Kosten anderer.	Take advantage of others.
		Mastery	
Ori	+	Ich bin ein kreativer Kopf.	Am an original thinker.
Per	+	Ich habe eine sehr reife Sichtweise auf das Leben.	Have a mature view on life.
Val	−	Ich stehe nicht für meine Überzeugungen ein.	Do not stand up for my beliefs.
Jud	+	Meine Freunde schätzen mich für mein gutes Urteilsvermögen.	Am valued by my friends for my good judgment.
Soc	−	Ich bin nicht gut darin, die Reaktionen anderer Leute vorherzusehen.	Have trouble guessing how others will react.
Lea	−	Ich bin nicht gut darin, andere zur Zusammenarbeit zu motivieren.	Have difficulty getting others to work together.

Note. IL = item label, K = item keying, CS = co-defining strength. Hop = hope, Zes = zest, Gra = gratitude, For = forgiveness, Cap = capacity for love, Pru = prudence, Mod = modesty/humility, Equ = equity, Int = integrity, Kin = kindness, Ori = originality, Per = perspective, Val = valor/bravery, Jud = judgment, Soc = social intelligence, Lea = leadership.

Results

Scale Development

In Table 3.3, we present the final 18-item solution in German and English that we chose to measure positivity, dependability, and mastery. In Table 3.4, we present their standardized factor loadings and factor correlations from the multi-group first-order factor model (including an acquiescence factor) in the training, validation, and use-case samples. In Table 3.5, we present the corresponding model fit indices, measurement invariance tests, and both the scales' composite reliabilities and test–retest correlations (i.e., in both countries). In Table S2a–c of the SOM, we present the zero-order correlations of the 18 items and their mean vectors in the training, validation, and use-case samples.

The final item selection measures each of the three core strengths with balanced-keyed scales and good construct representation. That is, all 14 marker strengths are represented in their respective core strength, two of them with two items (capacity for love and equity). In addition, two co-defining strengths are represented with one item each. Thereby, the two selected items from the co-defining strengths gratitude and kindness lent themselves well to operationalize the global constructs of positivity and dependability: The negatively keyed gratitude item ("Find few things in my life to be grateful for.") expresses a negative interpretation of situations or events—thus people scoring high on positivity tend to disagree with it. Likewise, the negatively keyed kindness item ("Get impatient when others talk to me about their problems.") indicates that a person is *not* a dependable, considerate contact for others—thus people scoring high on dependability tend to disagree with it.

Table 3.4

Standardized Factor Loadings and Factor Correlations

		Training samples (1st data collection)				Validation samples (2nd data collection)				Use-case samples (3rd data collection)			
		Germany (Sample 1)		UK (Sample 2)		Germany (Sample 3)		UK (Sample 4)		Germany (Sample 5)		UK (Sample 6)	
IL	K	Trait	ARS	Trait	ARS	Trait	ARS	Trait	ARS	Trait	ARS	Trait	ARS
Positivity (Pos)													
Hop	+	.60	.27	.52	.36	.62	.23	.57	.36	.67	.34	.60	.48
Zes	+	.61	.23	.58	.34	.61	.20	.61	.33	.70	.32	.63	.46
Gra (CS)	−	−.60	.22	−.55	.31	−.55	.19	−.50	.29	−.23	.27	−.20	.37
For	−	−.44	.21	−.44	.33	−.44	.18	−.47	.33	−.39	.30	−.35	.43
Cap1	−	−.60	.20	−.62	.32	−.52	.17	−.58	.32	−.59	.26	−.58	.41
Cap2	+	.47	.23	.43	.32	.45	.20	.44	.32	.50	.30	.45	.43
		.55	**.23**	**.52**	**.33**	**.53**	**.20**	**.53**	**.33**	**.51**	**.30**	**.47**	**.43**
Dependability (Dep)													
Pru	+	.33	.24	.36	.38	.30	.23	.30	.41	.39	.37	.45	.55
Mod	+	.39	.21	.35	.28	.28	.18	.26	.28	.48	.32	.45	.38
Equ1	+	.54	.24	.53	.33	.48	.21	.48	.36	.51	.33	.59	.48
Int	−	−.43	.22	−.46	.35	−.46	.20	−.45	.34	−.48	.29	−.59	.46
Kin (CS)	−	−.51	.22	−.58	.36	−.50	.20	−.49	.34	−.54	.29	−.65	.44
Equ2	−	−.55	.27	−.56	.39	−.51	.23	−.52	.40	−.63	.33	−.68	.45
		.46	**.23**	**.47**	**.35**	**.42**	**.21**	**.41**	**.35**	**.51**	**.32**	**.57**	**.46**
Mastery (Mas)													
Ori	+	.42	.22	.43	.34	.37	.18	.45	.37	.31	.30	.41	.47
Per	+	.52	.25	.50	.36	.55	.23	.57	.38	.47	.38	.52	.49
Val	−	−.42	.21	−.49	.37	−.42	.20	−.47	.35	−.40	.27	−.53	.43
Jud	+	.66	.27	.59	.37	.63	.25	.59	.37	.48	.40	.50	.49
Soc	−	−.49	.23	−.51	.36	−.43	.20	−.49	.36	−.48	.29	−.63	.45
Lea	−	−.50	.22	−.57	.38	−.53	.19	−.62	.35	−.50	.31	−.63	.46
		.50	**.23**	**.52**	**.36**	**.49**	**.21**	**.53**	**.36**	**.44**	**.32**	**.54**	**.47**
*r*Pos-Dep		.64		.69		.60		.61		.47		.43	
*r*Pos-Mas		.64		.77		.67		.78		.70		.79	
*r*Dep-Mas		.55		.78		.62		.72		.49		.61	
		.61		**.75**		**.63**		**.70**		**.55**		**.61**	

Note. IL = item label, K = item keying, ARS = acquiescent response style, CS = co-defining strength. Hop = hope, Zes = zest, Gra = gratitude, For = forgiveness, Cap = capacity for love, Pru = prudence, Mod = modesty/humility, Equ = equity, Int = integrity, Kin = kindness, Ori = originality, Per = perspective, Val = valor/bravery, Jud = judgment, Soc = social intelligence, Lea = leadership. Average absolute factor loadings and average factor correlations in bold face.

Table 3.5

Model and Scale Evaluation

Samples	Model fit			Measurement invariance test		Reliability			
				metric (ΔCFI/ΔRMSEA/ΔSRMR/ΔBIC)	scalar (ΔCFI/ΔRMSEA/ΔSRMR/ΔBIC)	ω (Pos/Dep/Mas)		r_{tt} (Pos/Dep/Mas)	
	CFI	RMSEA	SRMR			Germany	UK	Germany	UK
Training (1&2)	.903	.046	.057	−.006/.000/.008/−65.712	−.007/.001/.001/−59.497	.73/.63/.68	.73/.66/.73	.79/.73/.72[a]	.65/.73/.66[b]
Validation (3&4)	.899	.045	.063	−.014/.002/.014/−41.247	−.015/.002/.001/−37.981	.70/.58/.66	.74/.58/.74	.76/.63/.67[c]	.74/.60/.77[d]
Use-case (5&6)	.904	.042	.077	−.013/.002/.007/−51.403	−.021/.004/.003/−26.932	.66/.72/.63	.74/.81/.78	.72/.75/.72[e]	.80/.70/.68[f]

Note. CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, BIC = Bayesian Information Criterion, ω = McDonald's Omega, r_{tt} = test–retest correlations, Pos = Positivity, Dep = Dependability, Mas = Mastery. [a]$N_{(Pos/Dep/Mas)}$ = 117; [b]$N_{(Pos/Dep/Mas)}$ = 107; [c]$N_{(Pos/Dep/Mas)}$ = 111; [d]$N_{(Pos)}$ = 117, $N_{(Dep/Mas)}$ = 118; [e]$N_{(Pos/Dep)}$ = 139, $N_{(Mas)}$ = 126; [f]$N_{(Pos)}$ = 132, $N_{(Dep/Mas)}$ = 133.

Model and Scale Evaluation in the Training Samples

In the training samples (i.e., samples 1 and 2) used for item selection, measurement invariance tests based on ΔCFI, ΔRMSEA, ΔSRMR, and ΔBIC between both the configural and metric model and the metric and scalar model endorsed accepting scalar invariance. The scalar invariant measurement model itself showed good fit (i.e., the cutoffs for CFI, RMSEA, and SRMR were met; Table 3.5). Standardized loadings on the core strength factors ranged between $.33 \leq |\lambda| \leq .66$ in Germany (sample 1) and between $.35 \leq |\lambda| \leq .62$ in the UK (sample 2; Table 3.4). Standardized loadings on the acquiescence factor were on average .23 in Germany but .35 in the UK, which may reflect cultural differences in acquiescence (associated with different social norms, see Lechner et al., 2019). Factor correlations were on average .61 in Germany (ranging between .55–.64) and .75 in the UK (ranging between .69–.78), showing that the latent factors were related but discriminable. McDonald's ω ranged between .63–.73 in Germany and between .66–.73 in the UK, showing satisfactory composite reliability for heterogenous scales measuring global constructs.

Additionally, we computed r_{tt} coefficients based on a 2–3 weeks test–retest interval as important complementing information on the reliability of the scale scores, especially for short heterogeneous scales (e.g., McCrae et al., 2011). r_{tt} ranged between .72–.79 in Germany and between .65–.73 in the UK, respectively.

Model and Scale Evaluation in the Validation Samples

In the validation samples (i.e., samples 3 and 4), based on which ACO checks the degree of generality of the quality of its item selection, we largely replicated the findings based on the training samples. Overall, factor loadings (Table 3.4), model fit, measurement invariance, and reliability coefficients (Table 3.5) were similar to the findings in the training samples. There were only few minor exceptions: First, CFI of the scalar model (.899) and ΔCFI (−.014 and −.015 for metric and scalar measurement invariance, respectively) deviated slightly from the applied cutoffs and their values in the training samples. However, the deviation of the CFI was negligible. Furthermore, we can still accept scalar measurement invariance, because both ΔRMSEA and ΔSRMR did not exceed their critical cutoffs (Chen, 2007) and decreases in BIC from both the configural to the metric and from the metric to the scalar model were > 10. Second, the loadings on dependability of the prudence and the modesty/humility item in the validation samples (.26 ≤ λ ≤ .30) were slightly lower than in the training samples (.33 ≤ λ ≤ .39; see Table 3.4)—also leveling down reliability (both ω and r_{tt}) of the dependability scale in both Germany and the UK.

Model and Scale Evaluation in the Use-Case Samples

In the use-case samples (i.e., samples 5 and 6), to which we presented the final three 6-item core strength scales only (instead of all 96 IPIP-VIA-R items), we also largely replicated the good psychometric properties of the training samples. As in the validation samples, ΔCFI values were greater than in the training samples and exceeded the cutoff of −.010 (ΔCFI = −.013 and −.021 for metric and scalar measurement invariance, respectively, in the use-case samples; see Table 3.5), but again ΔRMSEA, ΔSRMR, and ΔBIC values clearly supported scalar measurement invariance. Unlike in the validation samples, the dependability scale showed good factor saturation in the use-case samples (i.e., .39 ≤ |λ| ≤ .68 across both countries and all reliability coefficients ≥ .70; see Tables 3.4 & 3.5). However, the gratitude item of the positivity scale had a small loading in the use-case samples (cross-country average λ = −.22). The reliability of the three scales was again good in the use-case samples: Apart from slightly lower ωs of the positivity scale and the mastery scale in Germany (compared to those in the training and validation samples), all ω and r_{tt} coefficients were (highly) satisfactory ranging between .68–.81. Furthermore, the factor correlations were on average lower in the use-case samples (.47 ≤ $r_{Germany}$ ≤ .70; .43 ≤ r_{UK} ≤ .79).

Scale Validation

We present the correlations between the three core strength scales and the various validation constructs and criteria in Table 3.6, including whether they supported our hypotheses. In addition, we visualized a summary of the results in Figure 3.2. In general, results were consistent across countries: Hypothesis testing yielded different results in Germany and the UK only in six out of 44 instances. Measured against Gignac and Szodorai's (2016) meta-analytic classification of small (r = .10), typical (r = .20), and relatively large (r = .30) effects, the effects we identified were substantial: Taking into account all statistically significant effects (i.e., p < .05), their average absolute size was $|\bar{r}|$ = .31 (ranging between .13 ≤ $|r|$ ≤ .76) in Germany and $|\bar{r}|$ = .35 (ranging between .14 ≤ $|r|$ ≤ .77) in the UK.

Looking at the nomological nets with Stability, Plasticity, social growth orientation, personal growth orientation, and general life satisfaction, all hypotheses were accurate except for the one regarding Stability and some regarding personal growth orientation: Contrary to expectations, Stability correlated most strongly with positivity and had the second-strongest correlations with both dependability and mastery, instead of showing the strongest positive correlation with dependability. In line with our expectations, Plasticity was most strongly positively correlated with mastery (expressing their common function to facilitate a flexible reaction to changes and integration of new information) and had the second-strongest positive correlation with positivity (expressing their common function to facilitate positive affectivity). In addition, Plasticity showed comparatively weak positive correlations with dependability. In Table S3 on OSF, we report the correlations between the three core strengths and the Big Five dimensions. In line with the Brunswikian symmetry principal, that suggests correlations to be strongest between constructs of similar abstractness (Nesselroade & McArdle, 1997; Wittmann, 1988), the correlations with the more specific Big Five were somewhat weaker than those with Stability and Plasticity but showed a similar pattern.

Table 3.6

Correlations Between Three Core Strength Scales and Validation Constructs and Criteria

		Positivity			Dependability			Mastery	
		Germany	UK		Germany	UK		Germany	UK
Stability		.76 ***	.77 ***	(++)	*.62* ***	*.58* ***		.56 ***	.60 ***
Plasticity	(+)	**.50** ***	**.55** ***		.21 **	.19 **	(++)	**.70** ***	***.61*** ***
Social Growth Orientation	(+)	**.29** ***	**.30** ***	(++)	**.74** ***	**.68** ***	(+)	**.22** **	**.24** ***
Personal Growth Orientation	(+)	*.22* **	*.12* ns	(−−)	**−.14** **	**−.19** **	(++)	*.17* *	**.20** **
Life Satisfaction	(++)	**.65** ***	**.66** ***		.22 *	.11 ns	(+)	**.32** **	**.43** ***
Self-enhancing behaviors *(power)*		−.15 **	−.25 ***	(−−)	**−.44** ***	**−.54** ***		−.25 ***	−.28 ***
Buying risky stocks.		.03 ns	−.23 ***	(−−)	−*.20* ns	**−.38** ***		−.06 ns	−.27 ***
Damaging other people's property.		−.31 **	−.36 ***	(−−)	**−.53** ***	**−.54** ***		−.30 **	−.33 ***
Intentionally engaging in a physical altercation.		−.20 **	−.24 **	(−−)	***−.38*** *	**−.53** ***		−.31 ***	−.31 ***
Taking a short-haul flight.		−.08 ns	.03 ns	(−−)	**−.32** **	**−.28** **		−.18 **	−.02 ns
Risk-taking behaviors *(hedonism & stimulation)*		−.09 ns	−.20 **	(−−)	**−.37** ***	**−.43** ***		−.18 *	−.23 ***
Using illegal drugs.		−.18 **	−.29 ***	(−−)	***−.23*** *	***−.30*** **		−.18 **	−.28 ***
Participating in gambling.		.08 ns	−.05 ns	(−−)	−*.11* ns	**−.23** ***		.02 ns	−.06 ns
Having changing sexual contacts within a month.		−.21 *	−.33 ***	(−−)	**−.52** ***	**−.44** ***		−.21 *	−.27 ***
Going on an adventure vacation.		.13 ns	.07 ns	(−−)	−*.16* ns	−*.16* ns		.01 ns	.00 ns
Using public transport without a valid ticket.		−.20 **	−.22 **	(−−)	**−.42** ***	**−.54** ***		−.29 ***	−.29 ***
Personal growth behaviors *(self-direction)*		.17 *	−.01 ns		−.06 ns	−.03 ns	(++)	*.21* ns	−*.05* ns
Developing a product or creating a piece of art.		.13 *	−.10 ns		−.06 ns	.00 ns	(++)	*.16* ns	−*.13* ns
Attending a continuing education class.		.17 *	.04 ns		−.04 ns	.01 ns	(++)	*.06* ns	−*.03* ns
Trying out novel technological devices or apps.		.09 ns	.06 ns		.03 ns	−.12 ns	(++)	**.25** **	*.04* ns
Activism and civil courage behaviors *(self-direction & universalism)*		.11 ns	.00 ns		−.06 ns	−.08 ns	(++)	*.08* ns	−*.03* ns
Signing a petition.		.03 ns	.02 ns		−.04 ns	.13 ns	(++)	−*.06* ns	−*.03* ns
Contacting a politician, government or local government official.		−.02 ns	−.06 ns		−.37 ***	−.29 ***	(++)	−*.09* ns	−*.18* *
Intervening when a stranger is treated unfairly.		.21 ns	.06 ns		.23 **	−.02 ns	(++)	***.28*** *	*.16* ns

(continues)

	Positivity			Dependability			Mastery	
	Germany	UK		Germany	UK		Germany	UK
Sustainable behaviors *(universalism)*	.24**	.15 ns	(++)	***.27*** **	**.33** ***	(++)	*.24* *	*.02* ns
Taking action to produce less waste.	.30***	.28***	(++)	**.42** ***	**.42** ***	(++)	***.31*** **	*.16* ns
Buying fair trade products.	.27**	.06 ns	(++)	***.31*** **	**.27** **	(++)	*.18* ns	*−.05* ns
Buying second-hand instead of new.	−.02 ns	−.04 ns	(++)	*−.04* ns	*−.03* ns	(++)	*.08* ns	*−.07* ns
Charitable behaviors *(universalism & benevolence)*	.23**	.14 ns	(++)	*−.04* ns	*−.02* ns		.00 ns	−.12 ns
Taking part in a charitable event or activity.	.16*	.08 ns	(++)	*−.11* ns	*−.10* ns		−.01 ns	−.10 ns
Donating money to those in need.	.24**	.17 ns	(++)	*.04* ns	*.09* ns		.01 ns	−.11 ns
Provision-making behaviors *(security)*	.08 ns	−.04 ns	(++)	*−.17* ns	*−.27* **		−.01 ns	−.14*
Seeing a doctor for a preventive or routine check-up.	.13 ns	.17 ns	(++)	*.04* ns	*.04* ns		.05 ns	−.01 ns
Acquiring a new capital investment.	.01 ns	−.21*	(++)	*−.31* *	*−.46* ***		−.12 ns	−.22***

Note. Row-wise hypothesized pattern of correlations between each validation construct/criterion and three VIA scales: (++) strongest positive correlation, (+) second-strongest positive correlation, (−−) strongest negative correlation. If two strongest or second-strongest correlations are hypothesized, we expected those correlations to be roughly equal in size. If no hypothesis is specified, no or a weaker correlation is expected than for hypothesized relationships with respective validation construct/criterion. Coefficients in bold face if hypothesis proved accurate. Coefficients in italics if hypothesis proved inaccurate. Coefficients in both bold face and italics indicate that the difference between the hypothesized and a non-hypothesized effect is $\Delta r < .10$. *** $p < .001$, ** $p < .010$, * $p < .050$, ns $p \geq .050$. We mention the supposable main driver/motivation of each behavioral cluster in parentheses.

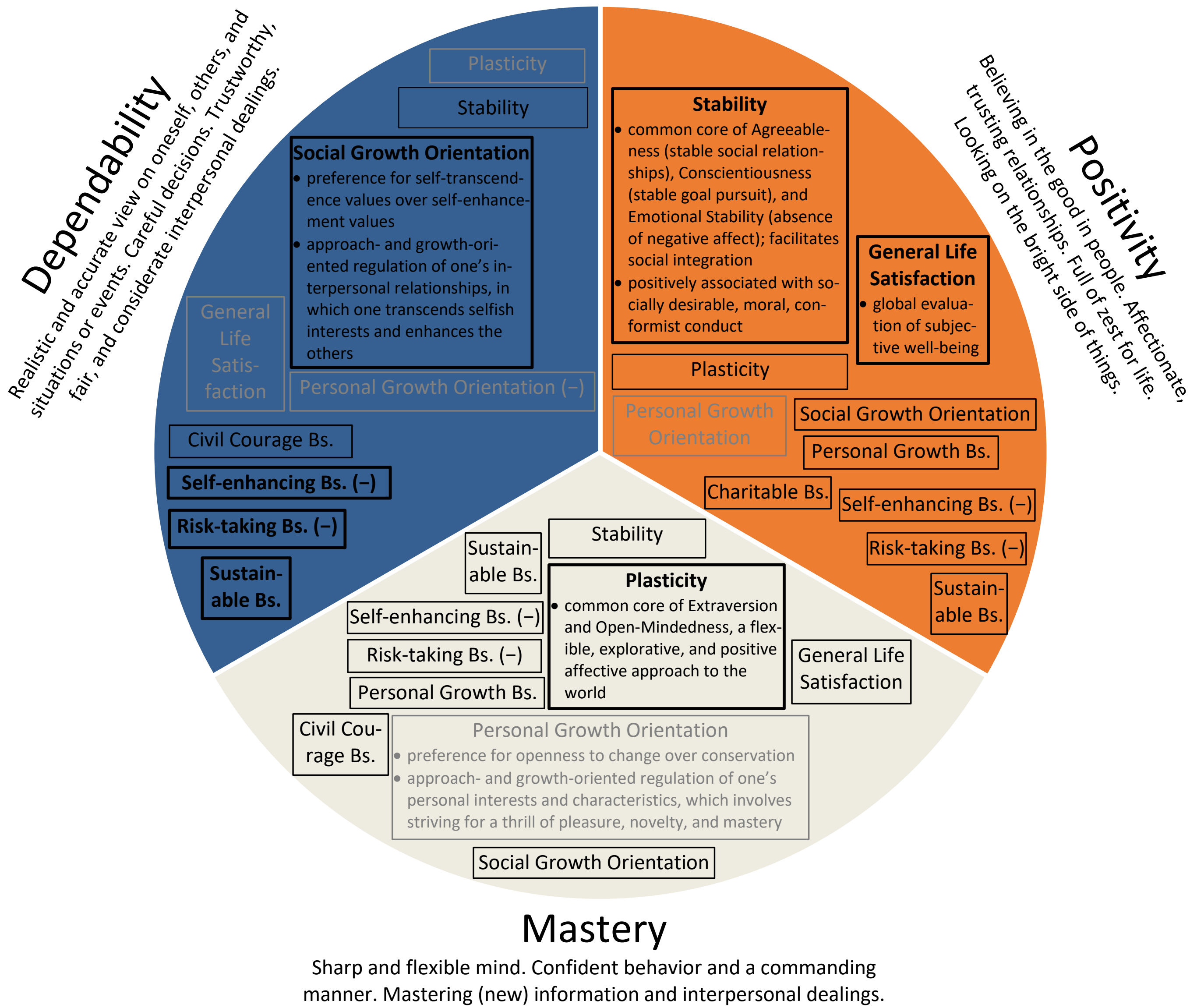


Figure 3.2. Nomological network of the three core strengths. We indicate in bold face, which of the three core strengths showed the strongest association with a validation construct/criterion. Correlates in gray font represent comparatively weak correlates within a core strength, providing evidence for discriminant validity. Correlates in black font provide evidence for convergent or criterion validity.

Regarding the higher-order values, social growth orientation was most strongly positively correlated with dependability and had the second-strongest positive correlation with both positivity and mastery (expressing the degree to which each core strength motivates to transcend selfish interests and enhance others). Personal growth orientation was most strongly negatively correlated with dependability in both countries (expressing the personal costs that come with the benefits of dependability for others). Furthermore, personal growth orientation was most strongly positively correlated with mastery in the UK (as expected) and with positivity in Germany (expressing the degree to which these core strengths involve a striving for mastery, novelty, or a thrill of pleasure). All correlations of the core strength scales with personal growth orientation were smaller than the corresponding correlations with social growth orientation. In line with our expectations, general life satisfaction was most strongly positively correlated with positivity and had the second-strongest positive correlation with mastery across countries. In addition, it showed a comparatively weak positive correlation with dependability in Germany.

The three core strength scales showed several of the expected and some unexpected correlations with the seven behavioral clusters and their items, respectively. The rather egocentric (1) self-enhancing and (2) risk-taking behaviors showed many negative correlations with all socially beneficial core strengths. In line with our expectations, those with dependability were strongest. (3) Only in Germany, personal growth behaviors were positively correlated with either positivity or, as expected, mastery (which may reflect the personal growth orientation expressed by both the core strengths and behaviors). (4) Only in Germany, the civil courage behavior “Intervening when a stranger is treated unfairly”, showed positive correlations with both dependability (likely attributable to the common *civil*-aspect) and, as expected, mastery (likely attributable to the common *courage*-aspect). Contrary to expectations, the two activism items were either not at all or negatively correlated with all core strengths in both countries. (5) Sustainable behaviors showed some positive correlations with all core strengths that, expectedly, tended to be strongest with dependability (which may reflect the social growth orientation expressed by both the core strengths and behaviors). (6) Charitable behaviors were positively correlated with positivity in Germany. The positivity facets gratitude and capacity for love may have driven these correlations (e.g., donating as reaction to gratitude for own privileges and awareness of being cared for). Furthermore, this finding is in line with Lavy and Benish-Weisman (2021) showing that gratitude mediates the relationship between self-transcendence values and prosocial behavior. Contrary to expectations, charitable behaviors were not correlated with dependability. (7) Finally, the expected positive correlations between provision-making behaviors and dependability were not confirmed. Instead, all correlations between provision-making behaviors and the core strength scales were either non-significant or negative.[11]

Discussion

Measurement of the VIA trait space has so far mainly focused on the 24 character strengths (Bluemke et al., 2021; McGrath et al., 2022), which have a similar level of abstraction as personality facets. By contrast, because of the longstanding lack of consensus regarding the number and nature of meaningful higher-order dimensions in the VIA framework, measures of factor-analytically determined global VIA traits have long been in short supply. The sole exception are the VIA-IS-V3 scales provided by the VIA Institute on Character upon registration that measure the three dimensions Caring, Inquisitiveness, and Self-Control (McGrath, 2019).

To allow for high-quality, open-access measurement of global VIA traits, we developed and validated scales measuring the three well-interpretable, truly global, and cross-culturally replicable VIA dimensions positivity, dependability, and mastery that Partsch et al. (2022) established in their factor-analytic study. For this purpose, we purposefully selected six items for each scale from a refined public domain VIA instrument, the 96-item IPIP-VIA-R (Bluemke et al., 2021), with the help of an ACO algorithm that simultaneously optimized multiple psychometric criteria. The resulting three core strength scales lend themselves to economically measure human character when global traits (as opposed to the 24 narrower VIA character strengths) are of interest.

The Core Strength Scales Show Good Psychometric Properties Across Samples

Using ACO, we selected three 6-item scales, that (1) represented the heterogeneity of the core strengths well, (2) were balanced-keyed to allow for acquiescence control by building unit-weighted manifest scale scores or modeling an acquiescence factor in a latent measurement model (Billiet & McClendon, 2000), (3) showed good model fit or factorial validity, (4) were scalar measurement invariant across Germany and the UK, allowing scale users to compare the latent means across countries. Further, the scales (5) represented distinct factors (i.e., factor correlations $r < .80$) that (6) showed good factor saturation (i.e., $|\lambda| > .30$) and (7) exhibited proper composite reliabilities as well as test–retest correlations of the manifest scale scores, especially given the high efficiency and heterogeneity of the scales (i.e., both ω and $r_{tt} > .60$), that were also comparable with reliability coefficients of scales measuring global constructs reported in the literature[12], allowing scale users to calculate manifest scale scores for analyses on the group level (i.e., research but not diagnostic purposes) as this is common practice (even though it is always advisable to use latent measurement models, especially for scales of heterogeneous constructs with comparatively much item uniqueness). Except for minor fluctuations, we reproduced these highly satisfactory psychometric properties in four independent samples collected one (i.e., the validation samples) and almost three years later (i.e., the use-case samples). In the use-case samples, we demonstrated that the newly developed scales also perform when administered in isolation in their final form outside of the original IPIP-VIA-R item pool.

Although the psychometric results were overall highly satisfying, two findings are in need of explanation. First, the weak loading of the gratitude item in the use-case samples begs the question if this item from a co-defining strength might not work in final scale use.

However, some sampling variation in factor loadings is, of course, expected and this particular finding might also be a temporary effect of the COVID-19 pandemic: It is reminiscent of a finding by Allemand et al. (2021), who showed lower correlations between gratitude and a positive outlook on one's own future in 2020 compared to the pre-pandemic year 2018. Second, the correlations between the three core strengths were substantial ($.43 \leq r \leq .79$ across all samples even when controlling for acquiescence). Although positive correlations between the three strengths are not unexpected, the size of these correlations might be inflated because we used only single-informant ratings. Previous research reported similar findings for personality metatrait scales. DeYoung (2006) and DeYoung et al. (2002) reported substantial correlations between latent Stability and Plasticity factors from CFA or SEM models based on single-informant ratings up to $r = .53$ that vanished in models controlling for specific rater variance (via a multi-informant approach; DeYoung, 2006). A multi-informant approach would likely also lower the factor correlations between the three core strengths. However, we would not expect them to completely vanish, because they likely arise from a *substantive* VIA g-factor (see also Partsch et al., 2022). As all VIA traits are socially desirable, people who generally adapt successfully to their (social) environment, tend to score higher on all three core strengths. Furthermore, people likely differ in their expectations regarding (good) character—while some impose generally low/mild standards when evaluating themselves (or others) by the core strength scales, others impose generally high/rigid standards.

Evidence for the Validity and Distinctness of the Three Core Strength Scales

Overall, the nomological nets with metatraits from personality and value frameworks, general life satisfaction, and a broad range of value-related behaviors supported the construct and criterion validity of each of the three scales. At the same time, correlations did not exceed .77, corroborating that the three core strengths are no replicas of the existing metatraits but distinct constructs. Furthermore, the distinct, meaningful correlation pattern for each scale also underscored the distinct nature of each of the three global VIA constructs. Most of the findings were robust across Germany and the UK. Deviations between countries appeared mainly in the nomological net with the behavioral correlates, whereby effects in the German sample were not replicated in the UK sample, attesting to the slightly stronger criterion validity of the three core strength scales in the German compared to the UK samples.

Regarding construct validity, Plasticity matched closest with mastery, whereas Stability—instead of matching closest with dependability—was strongly associated with all three core strengths. This suggests that all of them, but especially positivity, involve relationship, motivational and emotional stability to a substantial degree. DeYoung et al. (2002) showed that Stability—but not Plasticity—is positively associated with social desirability in terms of moral or conformist conduct. Thus, the high correlations between all three core strengths and Stability likely hail from their overlapping social and moral relevance, thereby underscoring the convergent validity of our scales.[13] The relatively small correlations between dependability and Plasticity provide evidence for discriminant validity. Furthermore, the correlation pattern with social and personal growth orientation showed both convergent and discriminant validity of the scales: Each core strength correlated more strongly with social growth orientation than with personal growth orientation, corroborating that all three core strengths are primarily *socially* beneficial. Whereas dependability as most socially relevant core strength even comes with small personal costs, positivity and mastery are also somewhat beneficial for the self. This is in line with the claim that "[a character] strength contributes to various fulfillments that constitute the good life, for oneself and for others. […]" (Peterson & Seligman, 2004; p. 17).

Regarding criterion validity, general life satisfaction/well-being is a crucial and therefore well-studied outcome of the VIA trait space. Our results are in line with previous research that showed positive correlations between life satisfaction and most or all of the 24 character strengths with hope, gratitude, capacity for love, and zest (i.e., character strengths covered by positivity) ranging among the strongest and modesty/humility and prudence (i.e., covered by dependability) ranging among the weakest correlates (Bluemke et al., 2021; Park et al., 2004; Proyer et al., 2011; Ruch et al., 2007; Weber et al., 2013). Furthermore, it is plausible that dependability as the most socially beneficial core strength, which also comes with some personal costs, is weakly associated with well-being at most.

Besides the cognitive appraisal of well-being, it is of particular interest how the three core strengths are related to actual (if only self-reported) behavior. Previous research suggests that basic human values are important drivers of behaviors and that VIA traits (i.e., "values in action") are mediators of these associations (Lavy & Benish-Weisman, 2021; Schwartz, 2015). Accordingly, we found several expected associations between the three core strengths and behaviors that were originally derived from different basic human values (see Table 3.6). In line with our findings regarding the higher-order values, the three core strengths showed most and on average strongest positive (negative) associations with *socially* beneficial (unfavorable) behaviors (i.e., self-enhancing, risk-taking, and sustainable behaviors). We also obtained some unexpected results that can be explained by the complex motivational basis of single behaviors, that is, single behaviors can be motivated/driven by different, sometimes even opposing values (Lins De Holanda Coelho et al., 2021; Schwartz, 2015). Thus, if correlations between the three core strengths and behaviors turned out differently than expected, this is most likely because the mechanism that we assumed behind a hypothesized correlation was not its *main* driver. For example, instead of being positively correlated with mastery, thereby expressing the intention to give good advice to political decision makers, the activism behavior "contacting a politician, government or local government official" was negatively correlated with dependability, potentially expressing more of a power or social dominance motive. Likewise, the provision-making behavior "acquiring a new capital investment" showed a negative instead of a positive correlation with dependability—obviously, this behavior was not primarily driven by a provision-making motive (i.e., the value of security) but a motive that is potentially negatively associated with dependability, for example a power motive. Moreover, behavior is not solely determined by motives, values, and traits but also by habits, automatisms, or situational aspects.

Limitations and Future Directions for Measuring the Three Core Strengths

Even though our ACO-based scale development approach yielded three core strength scales with highly satisfactory and robust psychometric properties and strong evidence for validity, our present study has limitations that future research should address. First, our focus here was on Germany and the UK. Future research should expand that focus by adapting the three scales for application in other, including non-WEIRD (i.e., white, educated, industrialized, rich, and democratic) cultures to test if scalar measurement invariance also holds across a more diverse set of countries and if the claim to universality of the 24 VIA character strengths can be generalized to the three global core strengths.

Second, researchers should apply the scales in multi-informant ratings to test if the psychometric properties generally replicate based on informant/observer ratings, how strongly self- and informant-rated core strengths are correlated (i.e., as rough indication of the role that socially desirable responding plays in the application of the scales; McCrae & Mõttus, 2019), and if the factor correlations decrease if latent core strengths variables are based on multi-informant data (as DeYoung, 2006, demonstrated for Stability and Plasticity).

Third, future research should further expand the nomological network of the three core strength scales beyond what was possible in the present study. In particular, it would be instructive to look at the correlations between the three core strength scales and the positivity construct by Caprara et al. (2012), Social Self-Regulation, Dynamism (i.e., the Big Two from the Big Six framework; Thalmayer & Saucier, 2014), and Caring, Inquisitiveness, and Self-Control from the three-virtue model (e.g., McGrath et al., 2018; McGrath & Wallace, 2021). Furthermore, it would be valuable to look at still more outcome associations, especially objective and informant-rated criteria to test criterion validity more rigorously.

Conclusion

Our study yielded three 6-item scales for the measurement of the global core strengths identified by Partsch et al. (2022): positivity, dependability, and mastery. We applied an ACO algorithm to optimize the item selection from the 96-item IPIP-VIA-R item pool in terms of model fit, scalar measurement invariance, factor saturation, and balanced keying to control for acquiescence. Beyond that, we ensured good construct coverage/representation in all scales and showed scale reliability. We replicated the psychometric properties of the scales across six samples from two countries and provided evidence for their construct and criterion validity. Thereby, we also showed that the three correlated core strengths are both meaningfully and distinguishably located in the nomological net of personality and value metatraits, well-being, and a broad range of behaviors. These highly performant scales, consisting of items from the IPIP, are the first for the measurement of global VIA traits that are accessible in the public domain. We hope they will stimulate both research advancing the understanding of the VIA trait space and applied character research warranting parsimonious measures of global character traits.

Endnotes

[1] Like Partsch et al. (2022), we refer to all abstract traits, that summarize at least three facet-level traits, as “global traits”. Following DeYoung et al. (2002), we refer to the two or three global traits on Levels II–III at the top of a trait hierarchy as “metatraits”.

[2] Note that our use of the terms “marker strengths” and “co-defining strengths” slightly deviates from their use in Partsch et al. (2022) who refer to all strengths loading $\lambda \geq .50$ ($.30 \leq |\lambda| < .50$) on a dimension in either one or both countries as marker strength (co-defining strength) of the respective dimensions in the respective country to interpret the results of the solutions-hierarchies revealed by the Bass-ackwards method. We refer to the 14 strengths loading $\lambda \geq .50$ on the same dimension across countries as marker strengths of the three core strengths; to the remaining 10 strengths, we refer to as co-defining strengths of the three core strengths.

[3] We excluded careless responders based on the Mahalanobis distance of the individual response vector from the mean sample response vector (Meade & Craig, 2012), the ipsatized variance across item scores (DeSimone & Harms, 2018), and the average response time per item (Leiner, 2019) in the main survey. In each sample from Germany and UK separately, respondents were flagged as careless responders if they fell within the upper 2.5% of the sample distribution of the Mahalanobis distance, or the lower 5% of the sample distribution of the ipsatized variance, or if their average response time per item was ≤ 1 second. Most estimates of the proportion of carless responders in a survey range between 5–15% (DeSimone & Harms, 2018). For example, Meade and Craig (2012) detected 10–12% careless responders in a student sample. Assuming that this can be considered as upper limit for our samples, we aimed for an exclusion rate below 10%.

[4] In the pre-registration we referred to these behaviors as “excitement-seeking behaviors” but renamed them when drafting the manuscript.

[5] We tried both more liberal and more restrictive approaches to achieve good construct coverage. In the more liberal, computationally intense approaches, we did not force ACO to choose items from each marker strength but rewarded the choice of items from many different (marker) strengths via optimization criteria for which we tried different weights. Apart from trials, in which we put high weights on the criterion optimizing construct coverage, these liberal approaches resulted in scales that were either biased towards specific character strengths or missed the definitional core of the three core strengths. In more restrictive approaches, we forced ACO to choose items not only from all 14 marker strengths but also from four different co-defining strengths. This resulted in too high factor correlations in the measurement model. Thus, we deem the chosen approach most suitable to achieve good construct coverage.

[6] As the factor, whose items are sampled first, has a higher selection probability of the co-defining items than a factor whose items are sampled afterwards, ACO chose the order in which it sampled items for each of the three factors randomly at each item selection.

[7] The recommendation of AVE $\geq .50$ by Fornell and Larcker (1981) is based on the assumption that the amount of variance *not* captured by the construct is error variance and should therefore be proportionally less. However, we assume that much of the variance *not* captured by the construct is rather specific item/facet uniqueness than random measurement error. Therefore, we deem it unproblematic if AVE falls (clearly) below .50.

[8] Our item sampling strategy required factor-specific pheromone updates for items from co-defining strengths (e.g., if the best selection of an iteration contained a gratitude item as indicator of positivity, then the pheromone level and thus the selection probability of this item was specifically increased for the positivity factor).

[9] As optimization criterion for ACO we chose AVE instead of McDonald's ω because we used non-recoded item scores in the ACO item selection for technical reasons. To compute McDonald's ω, item scores of negatively keyed items must be recoded. To evaluate the final

item selection, we used McDonald's ω to determine the composite reliability of the manifest scale scores as this is more relevant for future users of the scales.

[10] We also tried to include (low) factor correlations as ACO optimization criterion. However, then ACO did not manage to satisfy all optimization criteria with tolerable computational intensity (i.e., the automized item selection became too complex).

[11] As a side note, the Brunswikian symmetry effect of the behavioral clusters (i.e., a stronger correlation with the clusters than with the single items) failed to appear. The thematically clustered single behaviors were apparently too heterogeneous (i.e., too little correlated) to be aggregated on the same abstraction level as the global core strengths. Instead, quite unexpectedly, we found numerous effects on the item level with up to large effect sizes of $|r| > .50$.

[12] The composite reliabilities were comparable with the ωs that Rammstedt et al. (2021) reported for the somewhat more homogenous 6-item Big Five dimension scales of the BFI-2-S (i.e., $.52 \leq \omega \leq .83$) and at least as high as the internal consistencies that researchers reported for scales of comparably heterogenous constructs (e.g., $.54 \leq \alpha \leq .63$ for the 7-item Social Self-Regulation scale and $.51 \leq \alpha \leq .62$ for the 7-item Dynamism scale; Thalmayer & Saucier, 2014). The test–retest correlations of the three core strength scales were somewhat smaller than those reported for the BFI-2-S domain scales (.79–.88 after six weeks, Rammstedt et al., 2020; .69–.83 after 2 months, .77–.88 after 3 months, Soto & John, 2017b), but comparable in size with those reported by Caprara et al. (2012) for their 8-item positivity scale ($.69 \leq r_{tt} \leq .73$ after six weeks).

[13] Partsch et al. (2022) also assumed that Social Self-Regulation from the Big Six framework (i.e., the common core of Agreeableness, Conscientiousness, and Honesty/Propriety; Thalmayer & Saucier, 2014) might be the most closely related personality metatrait of dependability (rather than Stability).

GENERAL DISCUSSION

Main Findings and Contributions of the Dissertation in the Light of Its Research Aims

The dissertation defined its research aims against the backdrop of several unresolved methodological challenges in the assessment of VIA character as well as in the examination of its structure or trait hierarchy. Firstly, while the VIA Institute on Character recently refined its proprietary instruments for character assessment based on the VIA framework, the only public domain instrument for the assessment of the 24 character strengths, the IPIP-VIA, still suffers from its teething problems almost 20 years later. Secondly, due to the years of general standstill in VIA scale development and great variety in methodological approaches, factor-analytic studies exploring the hierarchical structure of the 24 VIA character strengths yielded inconsistent results over the years and only recently converged to the finding that three global dimensions span the character trait space. Thirdly, due to the fact that the VIA trait hierarchy and its global dimensions have not been conclusively established, there was only one proprietary measure of global VIA traits available—and none in the public domain. When meeting these methodological challenges, the dissertation constantly put a strong emphasis on practicing open science, showing replicability of findings (among others, across countries to pave the way for cross-cultural character research), and integrating findings into the bigger scope of personality science.

Study 1 yielded the public domain IPIP-VIA-R for the measurement of the 24 VIA character strengths in both English and German. The IPIP-VIA-R exclusively relies on field-tested items from the public domain IPIP. A team of experts from the fields of psychometry, personality psychology, cross-cultural research, and translation studies selected the English-language items and adapted them to German. The balanced-keyed, content-valid, mostly unidimensional, and partially or fully scalar measurement invariant (i.e., across the UK and Germany) 4-item scales remedy pressing psychometric issues of the IPIP-VIA and are particularly suitable for application in large-scale surveys. Complementing previous research by McGrath et al. (2020) and Ruch et al. (2021) (see Table 1 of the dissertation's General Introduction), Study 1 showed likely redundancy for 4 (7) VIA strengths with Big Five facets in the UK (Germany) by using the BFI-2-S instead of a NEO inventory and applying the cut-off for "likely redundancy" suggested by McGrath et al. (2020) (i.e., an uncorrected correlation between a character strength and a Big Five facet of $|r| \geq .60$). Thus, Study 1 supported previous findings that most character strengths can be considered distinct from personality traits.

Based on the public domain IPIP-VIA-R established in Study 1, Study 2 revisited the hierarchical structure of the 24 character strengths. Results revealed three well-interpretable, global *core strengths* that replicated across Germany and the UK (i.e., *positivity*, *dependability*, and *mastery*), corroborating recent evidence that the VIA trait space is spanned by three global dimensions. To integrate the core strengths into the scope of personality science, Study 2 drew on construct definitions of several individual differences constructs and related *positivity* to Caprara et al.'s (2012) positivity construct, *dependability* to Social Self-Regulation (i.e., metatrait from the Big Six framework; furthermore also to the metatrait Stability from the Big Five framework as well as to communion from Bakan's (1966) duality of human existence), and mastery to Dynamism (furthermore also to Plasticity and agency).

Finally, Study 3 yielded an 18-item subset of the IPIP-VIA-R to measure each core strength economically with a balanced-keyed, reliable, and (content-, construct-, and criterion) valid 6-item scale that showed scalar measurement invariance across Germany and the UK. Nomological network analyses empirically supported the theoretical association between mastery and Plasticity established in Study 2, while Stability was unexpectedly more strongly related to positivity than to dependability.

Broader Methodological Contributions

Meeting three methodological challenges in the context of character assessment, the dissertation made some methodological contributions that also can be applied to other fields of individual differences research. Study 1 demonstrated how researchers can proceed to refine personality scales within the IPIP (i.e., the overarching goal of the public domain platform; Goldberg et al., 2006) and translate/adapt them to other languages. Thereby, the rational-empirical approach demonstrated that statistical criteria for item selection (e.g., reliability) can be fruitfully complemented by expert ratings of content validity and cross-cultural comparability in order to optimize the scales along multiple dimensions of quality. Even though Study 1 pointed out potential limitations of the IPIP for scale development or refinement (e.g., the limited availability of negatively keyed items makes it hard to construct balanced-keyed scales), survey researchers might adopt the presented approach to other IPIP instruments. For example, if survey programs require short scales to measure vocational interests, they might use the eight 10–14-item Oregon Vocational Interest Scales (ORVIS; Pozzebon et al., 2010) to select the most content-valid items applicable to the target population of the survey and complement them with suitable items from other IPIP scales (e.g., items from the NEO excitement-seeking scale for the ORVIS adventure short scale).

Study 2 demonstrated how researchers can proceed to identify *useful* global constructs in a trait hierarchy. Toward that end, Study 2 performed a factor-analytic triad of parallel analysis/MAP analysis, Bass-ackwards analysis, and congruency analysis and introduced three criteria that must apply to each dimension of a useful global level in a solutions-hierarchy: *interpretability* (i.e., unique highly loading marker variables); *globality/generality* (i.e., at least three marker variables ensuring that a dimension aggregates a significant amount of specific variation); and *cross-cultural replicability* (i.e., high component or factor congruency across countries). The criterion of cross-cultural replicability serves not only the identification of cross-culturally relevant traits but also evidences the robustness of the revealed dimensions (i.e., guards against sample-specific findings). For the latter purpose, researchers could show factor congruency also across different samples from the same country. Researchers may apply the three criteria to clarify the hierarchical structure of any other construct space. For example, in the light of their overlap outlined in the General Introduction of this dissertation, it would be interesting to synthesize the VIA, the Big Five, and the Big Six framework in a factor-analytic study by using the IPIP-VIA-R as well as facet scales of different Big Five and Big Six questionnaires (e.g., the BFI-2, the NEO-PI-R, and the HEXACO-PI-R; Ashton & Lee, 2008; Costa & McCrae, 1992; Soto & John, 2017a), conducting the factor-analytic triad from Study 2, and interpreting the resulting solutions-hierarchies based on the three criteria. How many—and which—useful global dimensions could be extracted? Which of the already known global personality and character traits would be revealed separately, and which would blend?

Study 3 demonstrated how researchers can proceed to develop scales of newly identified global traits based on an algorithmic item selection approach. Hitherto, researchers applied metaheuristic algorithms like an ACO or a genetic algorithm to develop short scales based on their longer version (Schroeders et al., 2016). The development of (short) scales measuring newly identified global constructs has two additional requirements: First, not all items of the item pool can be clearly assigned to be selected for one specific scale (i.e., in Study 3, these were the items measuring co-defining strengths). To meet this requirement, Study 3 applied a newly programmed item sampling mechanism for ACO. Second, even though this newly introduced item sampling mechanism considers good construct coverage, not all (best) solutions that ACO suggests based on the optimization criteria are ideal in terms of content validity. Therefore, Study 3 suggested a hybrid approach to item selection: Instead of automatically accepting the overall best solution resulting from several ACO runs (i.e., the method of choice for ACO-based short scale development), experts evaluated all best solutions from different ACO runs in terms of content validity to ultimately choose the global trait scales. Thus, like Study 1, Study 3 used a rational-empirical approach to scale development that complemented statistical criteria with expert ratings of content validity.

Researchers can follow this approach to develop scales for any newly identified individual differences constructs (and re-use our openly accessible ACO R code to that end).

Limitations and Directions for Future Research

In this section, I point out three main limitations that all studies of the dissertation share. I then derive implications for future research in the subsequent subsections. First, the dissertation based all studies on samples from two culturally similar countries—Germany and the UK (plus the US convenience sample from Study 1.1). Even though this was a reasonable first step towards a cross-culturally valid VIA trait hierarchy and cross-culturally adaptable character and core strength scales, and it supported the robustness of our results, the dissertation did not show real cross-cultural replicability. Thus, it remains unclear (a) if the (partial) scalar measurement invariance of the newly developed character and core strength scales would also hold across culturally diverse, non-WEIRD (i.e., white, educated, industrialized, rich, and democratic) countries, (b) if—at least metric measurement invariance provided—the nomological nets (e.g., with basic human values) and criterion correlations (e.g., with life satisfaction) would be comparable with those from Germany and the UK to support the universality of the positively valued nature of (VIA) character traits, and (c) if the factor-analytic approach from Study 2 would reveal the three core strengths (i.e., positivity, dependability, and mastery) as well-interpretable, truly global character traits replicable across diverse cultures in the first place.

Second, the dissertation based all studies on single-informant data (collected with self-reports). Even though this is the usual first step in scale development and validation as well as in the analysis of a construct's hierarchical structure, single-informant-based findings are especially limited in the context of character constructs because of their strong evaluative nature and the socially desirable responding involved that potentially results in overestimated correlations between the scales. Thus, despite the well-founded assumption outlined in Study 2 that social desirability is not a mere response style but also involves actual socially adaptable behavior observable by others (i.e., true variance), it remains unclear if the trait hierarchy established in Study 2 as well as the findings on construct- and criterion validity of the IPIP-VIA-R character and core strengths from Studies 1 and 3 fully replicate when using multi-informant scale scores purged of evaluative biases (i.e., idiosyncratic source method variance; McCrae & Mõttus, 2019).

Third, Studies 1 and 3 put the character and core strengths (scales) in a nomological net with Big Five personality constructs only. Despite the Big Five's popularity and prevalence in describing personality, the dissertation's General Introduction outlined that it is even more instructive to compare VIA traits with Big Six (or "HEXACO") traits. Thus, the dissertation only partially showed the empirical distinctness versus redundancy between character as measured by the IPIP-VIA-R and personality.

Future Directions for Character Assessment

The limitations of this dissertation prompt implications for future research with the IPIP-VIA-R character and core strength scales. First, to show real cross-cultural replicability and universality of both the constructs and the scales measuring them, researchers need to translate and adapt the cross-culturally adaptable items of the IPIP-VIA-R to languages from a diverse set of cultures and both test and validate the scales based on preferably representative samples. I recommend following the approach that Study 1 used to obtain the German version of the IPIP-VIA-R: To translate or adapt the items according to the TRAPD approach introduced by Harkness (2003) and document the process and decisions as described by Behr and Zabal (2021). With the adaptions being made in the German version, it turned out to be psychometrically somewhat superior to the English source version and thereby qualified itself as the source version for further translations. Alternatively, future translation projects should at least consider the documentation of the adaptions provided in English as supplementary material of Study 1 (https://osf.io/yz87n/).

Second, future research should extend the research presented in this dissertation to include a multi-informant approach. This would help to minimize idiosyncratic source method biases by single raters, such as socially desirable responding (McCrae & Mõttus, 2019). Revealing the core strengths of *positivity*, *dependability*, and *mastery* by rerunning the factor-analytic triad of Study 2 with multi-informant-based scores of the IPIP-VIA-R character strength scales would strongly corroborate their relevance. Furthermore, (re-)building the nomological net of VIA character traits and both Big Five and Big Six personality traits with multi-informant-based scale scores might show that character and personality frameworks are more distinct (i.e., less redundant) than the current state of research using single-informant-based scores suggests (i.e., Studies 1 and 3 and previous research reviewed in the General Introduction of the dissertation). One way to obtain multi-informant-based trait scores would be to factor the scores of each scale from self-ratings and the ratings by several informants such as spouses, friends, and colleagues (i.e., to extract the common, true variance across all informants) and save the resulting factor scores.

Third, future research should extend the construct and criterion validation of the 24 IPIP-VIA-R character strength scales. While Studies 1.6 and 1.7 attempted to locate the character strengths in the personality and value trait space as well as to test their explanatory power for two general character-related outcomes (i.e., general life satisfaction and subjective health), a more specific validation approach is pending. Ideally, researchers will conduct a large-scale validation study including many specific validation variables to predict and test a differentiated nomological network. For a more specific test of convergent validity, the study could include a scale of each character trait that was developed outside of the VIA framework (e.g., the GQ-6 scale of McCullough et al., 2002, as a measure of dispositional gratitude). For a more specific test of criterion validity, the study could use a large set of behavioral items (e.g., those applied in Study 3 or those from Bardi & Schwartz, 2003) or, ideally, develop behavioral markers for each character strength (e.g., "How often do you cut the acquaintance with somebody you previously considered a friend?" as a behavioral marker of (low) forgiveness). To evidence the IPIP-VIA-R scales' discriminant validity from each other, the study should show distinct correlation profiles across these various validation variables.

Finally, the refined IPIP-VIA-R comprises short and (whenever possible) unidimensional scales. They are designed for survey research and suitable to assess the core of each strength, while they are *not* suitable for individual diagnostics or a multi-faceted assessment of character strengths. For example, Fowers et al. (2021) suggested that character-related knowledge, motivation, disposition, and behavior are distinguishable and have incremental explanatory value beyond each other. Likewise, Goodman et al. (2019) suggested distinguishing if a person possesses a character trait, if they are aware of its possession, and if they use it. They argue that this would allow us to better understand *how* character trait effects are achieved. Furthermore, although Peterson and Seligman (2004) emphasized the "family resemblance" or common core among the synonyms that they used to describe each character strength, these synonym lists might also allow for multi-faceted construct operationalization. For example, *hope* comes with the synonyms or non-redundant facets of optimism, future-mindedness, and future orientation, or *curiosity* comes with interest, novelty-seeking, and openness to experience. Thus, future research should also develop refined longer, multi-faceted versions of VIA character strength scales that allow for both more differentiated and reliable assessment in diagnostic contexts (e.g., clinical, educational, or personnel development). To develop such longer scales for the public domain, researchers will have to generate new, tailored item material, since Study 1 demonstrated the challenge of identifying only four suitable items within the IPIP to operationalize some of the character strengths. Clearly, including such newly developed VIA long scales into the IPIP would enrich the same.

Such longer scales would also be (more) suitable to test various of the STRIVE-4 model hypotheses. STRIVE-4 was only recently introduced by Fowers et al. (2021) as a unifying framework for extant and future character research. It suggests 26 concise and testable hypotheses on the nature of character constructs that partly recover Peterson and Seligman's (2004) ten definitional criteria for character strengths.

Future Directions for Research on Character and Its Relationship to Personality

Previous research examining the distinctness versus redundancy between the character and personality trait space (McGrath et al., 2020; Ruch et al., 2021) built on either the original VIA-IS or its open-access cousin, the IPIP-VIA. Given the rudimentary development of these VIA inventories, the important question of empirical distinctness versus redundancy needs to be revisited based on revised character strength scales. Therefore, future research should complete the joint nomological net of IPIP-VIA-R-based character traits and personality traits by including constructs from the Big Six framework, that is, HEXACO facets and the metatraits Dynamism and Social Self-Regulation. Furthermore, the BFI facets or BFI-based Plasticity and Stability from Studies 1 and 3 could be complemented by NEO measures and the newly published NB5I scales that measure the Big Five with evaluatively neutralized items and are ideal for research that "focuses on discriminant validity, such as identifying theoretically valid relationships between personality traits and other concepts" (Bäckström et al., 2023; Abstract).

In addition to the nomological network approach, further approaches to integrate the VIA, Big Five, and Big Six frameworks would be instructive. A group of researchers relatively recently introduced a new framework for the integration of a wide range of personality constructs: the Circumplex of Personality Metatraits (CPM) that models personality with four bipolar metatraits building an octant structure in a two-dimensional space (e.g., Strus et al., 2014; Strus & Cieciuch, 2017). In its recently refined version (Strus & Cieciuch, 2021), the CPM is spanned by two orthogonal CPM metatraits that merge the Big Two metatraits from the Big Five and the Big Six framework: Stability and Social Self-Regulation are merged into Alpha+, while Plasticity and Dynamism are merged into Beta+.

Their newly introduced opposite poles are Alpha− (Disinhibition) and Beta− (Passiveness), respectively. To yield its octant structure, the CPM located the newly introduced metatrait Gamma+ (Integration) equidistantly between Alpha+ and Beta+, while locating the newly introduced metatrait Delta+ (Self-Restraint) equidistantly between Alpha+ and Beta−. Their opposite poles are Gamma− (Disharmony) and Delta− (Sensation-Seeking), respectively. For a visual representation of the CPM, see Figure 2 in Strus and Cieciuch (2021).

Following the approach by Strus and Cieciuch (2019), future research could empirically locate the VIA metatraits based on the scales of Study 3 in the refined CPM. Of course, this presumes that the three core strengths can be represented in a two-dimensional space (i.e., a crucial part of their co-variance must be explained by two orthogonal factors). Given their conception, intercorrelated nature, and their correlations with the Big Five, reported in the supplement of Study 3 (https://osf.io/wzk5t), one might expect that all three core strengths would locate within the quadrant marked out by Alpha+ and Beta+. Having Gamma+ midway, this quadrant "brings together all socially and individually desirable qualities of personality" (Strus & Cieciuch, 2019; p. 21)—thus, by definition, character traits should gather there. Presumably, based on the reasoning and empirical evidence from Studies 2 and 3, positivity would locate near Gamma+ (Integration), dependability would locate near Alpha+ (Stability/Social Self-Regulation), and mastery would locate near Beta+ (Plasticity/ Dynamism). As outlined above, the analyses suggested that future research on the integration of character and personality would be most instructive if based on a multi-informant approach.

Finally, the psychometric work in this dissertation is not an end in itself. Rather, the IPIP-VIA-R character and core strength scales enable substantive research on determinants and outcomes of character traits from various basic and applied areas of psychology. For example, are there neural or genetic correlates/determinants of character traits? Or how do character traits perform in predicting academic and career success (beyond personality and intelligence)? Furthermore, provided that the scales demonstrate sensitivity to change, they can be used to study the developmental trajectories of character traits across the lifespan.

General Conclusion

To conclude, the present dissertation advanced VIA character measurement in the public domain as well as the understanding of the VIA character trait space and hierarchy. The 96-item IPIP-VIA-R for the measurement of the 24 character strengths and its 18-item subset for the measurement of the three core strengths provide personality researchers with rigorously developed and openly accessible both facet- and global level character trait measures: Developed in parallel in English and German, the scales are economical, as reliable as can be expected of short scales, valid, comparable (i.e., replicable) across Germany and the UK, and adaptable for use in other countries and cultures. Based on the new scales, the dissertation contributed to the integration of the VIA character constructs into the bigger picture of personality science by relating the character and core strengths to (or predicting them by) Big Five and value constructs. Using the IPIP-VIA-R, the dissertation also contributed to the establishment of the VIA character trait hierarchy. Framing the research question of each study as a methodological challenge, the dissertation also made some general methodological contributions that researchers on individual differences might consider valuable. Overall, the dissertation accomplished its aims—and pointed out several more challenges for future research to take up. While having been largely neglected by personality researchers during the last century, character traits have been on the rise since the turn of the millennium. Considered beneficial by human beings across time, *good character* might also be indispensable to meet the global challenges of the 21st century.

REFERENCES

Ackerman, C. E., Warren, M. A., & Donaldson, S. I. (2018). Scaling the heights of positive psychology: A systematic review of measurement scales. *International Journal of Wellbeing*, *8*(2), 1–21. https://doi.org/10.5502/ijw.v8i2.734

Adams, R. M. (2006). *A theory of virtue: Excellence in being for the good*. Oxford University Press.

Adelnia, A., & Vahid Dastjerdi, H. (2011). Translation of idioms: A hard task for the translator. *Theory and Practice in Language Studies*, *1*(7), 879–883. https://doi.org/10.4304/tpls.1.7.879-883

Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality*, *53*, 1–4. https://doi.org/10.1016/j.jrp.2014.07.001

Aichholzer, J. (2015). Controlling acquiescence bias in measurement invariance tests. *Psihologija*, *48*(4), 409–429. https://doi.org/10.2298/PSI1504409A

Allalouf, A., Hambleton, R. K., & Sireci, S. G. (1999). Identifying the causes of DIF in translated verbal items. *Journal of Educational Measurement*, *36*(3), 185–198. https://doi.org/10.1111/j.1745-3984.1999.tb00553.x

Allemand, M., Olaru, G., & Hill, P. L. (2021). Gratitude and future time perspective during the COVID-19 pandemic: Effects of age and virus worry. *The Journal of Positive Psychology*. Advance online publication. https://doi.org/10.1080/17439760.2021.1975156

Allport, G. W. (1927). Concepts of trait and personality. *Psychological Bulletin*, *24*(5), 284–293. https://doi.org/10.1037/h0073629

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*(1), i–171. https://doi.org/10.1037/h0093360

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). American Psychiatric Association.

Anjum, A., & Amjad, N. (2019). Values in action inventory of strengths: Development and validation of short form-72 in Urdu. *Current Psychology*. Advance online publication. https://doi.org/10.1007/s12144-018-0102-6

Ashton, M. C. (2005, January). How the IPIP benefits personality research and education. In L. R. Goldberg (Chair), *The international personality item pool and the future of public-domain personality measures*. Presidential symposium at the 6th annual meeting of the Association for Research in Personality (ARP), New Orleans.

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

Ashton, M. C., & Lee, K. (2008). The prediction of Honesty–Humility-related criteria by the HEXACO and Five-Factor Models of personality. *Journal of Research in Personality, 42*(5), 1216–1228. https://doi.org/10.1016/j.jrp.2008.03.006

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO model of personality structure—And why those objections fail. *European Journal of Personality, 34*(4), 492–510. https://doi.org/10.1002/per.2242

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Blas, L. D., Boies, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*(2), 356–366. https://doi.org/10.1037/0022-3514.86.2.356

Asparouhov, T., & Muthén, B. O. (2018). *SRMR in Mplus*. http://www.statmodel.com/download/SRMR2.pdf

Azañedo, C. M., Fernández-Abascal, E. G., & Barraca, J. (2014). Character strengths in Spain: Validation of the Values in Action Inventory of Strengths (VIA-IS) in a Spanish sample. *Clínica y Salud*, *25*(2), 123–130. https://doi.org/10.1016/j.clysa.2014.06.002

Bäckström, M., Björklund, F., Maddux, R. E., & Lindén, M. (2023). The NB5I: A full-scale Big-Five inventory with evaluatively neutralized items. *European Journal of Psychological Assessment*. Advance online publication. https://doi.org/10.1027/1015-5759/a000687

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. Rand Mcnally.

Bardi, A., & Schwartz, S. H. (2003). Values and behavior: Strength and structure of relations. *Personality and Social Psychology Bulletin*, *29*(10), 1207–1220. https://doi.org/10.1177/0146167203254602

Beauducel, A. (2011). Indeterminacy of factor score estimates in slightly misspecified confirmatory factor models. *Journal of Modern Applied Statistical Methods*, *10*(2), 583–598. https://doi.org/10.22237/jmasm/1320120900

Beauducel, A., & Leue, A. (2014). Testing the assumption of uncorrelated errors for short scales by means of structural equation modeling. *Journal of Individual Differences*, *34*(4), 201–211. https://doi.org/10.1027/1614-0001/a000135

Béguin, C., & Hulliger, B. (2004). Multivariate outlier detection in incomplete survey data: The epidemic algorithm and transformed rank correlations. *Journal of the Royal Statistical Society, A 167*(Part 2.), 275–294. https://doi.org/10.1046/j.1467-985x.2003.00753.x

Behr, D., & Zabal, A. (2021). *Documenting Survey Translation* [GESIS - Survey Guidelines]. GESIS - Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_035

Beierlein, C., Kovaleva, A., László, Z., Kemper, C. J., & Rammstedt, B. (2015). *Kurzskala zur Erfassung der Allgemeinen Lebenszufriedenheit (L-1)*. Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/ZIS229

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences*, *42*(5), 825–829. https://doi.org/10.1016/j.paid.2006.09.024

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

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, *65*(5), 676–696. https://doi.org/10.1177/0013164404272507

Berry, J. W. (1969). On cross-cultural comparability. *International Journal of Psychology, 4*(2), 119–128. https://doi.org/10.1080/00207596908247261

Berry, J. W. (1999). Emics and etics: A symbiotic conception. *Culture and Psychology, 5*(2), 165–171. https://doi.org/10.1177/1354067X9952004

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, *7*(4), 608–628. https://doi.org/10.1207/S15328007SEM0704_5

Bilsky, W., Janik, M., & Schwartz, S. H. (2010). The structural organization of human values: Evidence from three rounds of the European Social Survey (ESS). *Journal of Cross-Cultural Psychology*, *42*(5), 759–776. https://doi.org/10.1177/0022022110362757

Biswas-Diener, R. (2006). From the equator to the North Pole: A study of character strengths. *Journal of Happiness Studies*, *7*(3), 293–310. https://doi.org/10.1007/s10902-005-3646-8

Bluemke, M., Jong, J., Grevenstein, D., Mikloušić, I., & Halberstadt, J. (2016). Measuring cross-cultural supernatural beliefs with self-and peer-reports. *PLoS ONE*, *11*(10), Article e0164291. https://doi.org/10.1371/journal.pone.0164291

Bluemke, M., Partsch, M. V., Saucier, G., & Lechner, C. M. (2021). *Human character in the IPIP: Towards shorter, more content-valid, and cross-culturally comparable IPIP-VIA character strength scales*. PsyArXiv. https://doi.org/10.31234/osf.io/k79qf

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Bollen, K. A., & Diamantopoulos, A. (2017). In defense of causal-formative indicators: A minority report. *Psychological Methods*, *22*(3), 581–596. https://doi.org/10.1037/met0000056

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Erlbaum.

Brdar, I., & Kashdan, T. B. (2010). Character strengths and well-being in Croatia: An empirical investigation of structure and correlates. *Journal of Research in Personality*, *44*(1), 151–154. https://doi.org/10.1016/j.jrp.2009.12.001

Brosseau-Liard, P. E., & Savalei, V. (2014). Adjusting relative fit indices for nonnormality. *Multivariate Behavioral Research*, *49*(5), 460–470. https://doi.org/10.1080/00273171.2014.933697

Brosseau-Liard, P. E., Savalei, V., & Li, L. (2012). An investigation of the sample performance of two non-normality corrections for RMSEA. *Multivariate Behavioral Research*, *47*(6), 904–930. https://doi.org/10.1080/00273171.2012.715252

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80*(4), 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, *27*(4), 509–540. https://doi.org/10.1207/s15327906mbr2704_2

Byrne, B. M. (1998). *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press.

Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (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

Cambré, B., Welkenhuysen-Gybels, J., & Billiet, J. (2002). Is it content or style? An evaluation of two competitive measurement models applied to a balanced set of ethnocentrism items. *International Journal of Comparative Sociology*, *43*(1), 1–20. https://doi.org/10.1177/002071520204300101

Caprara, G. V., Alessandri, G., Eisenberg, N., Kupfer, A., Steca, P., Caprara, M. G., Yamaguchi, S., Fukuzawa, A., & Abela, J. (2012). The positivity scale. *Psychological Assessment*, *24*(3), 701–712. https://doi.org/10.1037/a0026681

Castro Solano, A., & Cosentino, A. C. (2018). IVyF abreviado —IVyFabre—: Análisis psicométrico y de estructura factorial en Argentina [sci brief form (sci-bf): Psychometric and factor structure analyses in Argentina]. *Avances en Psicología Latinoamericana*, *36*(3), 619–637. http://dx.doi.org/10.12804/revistas.urosario.edu.co/apl/a.4681

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*(3), 153–193. https://doi.org/10.1037/h0059973

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://doi.org/10.1037/h0046743

Cattell, R. B. (1986). The psychometric properties of tests: Consistency, validity, and efficiency. In R. B. Cattell & R. C. Johnson (Eds.), *Functional psychological testing* (pp. 54–78). Brunner/Mazel.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *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

Chernyshenko, O. S., Kankaraš, M., & Drasgow, F. (2018). *Social and emotional skills for student success and well-being: Conceptual framework for the OECD study on social and emotional skills* (OECD Education Working Papers No. 173). OECD Publishing. https://doi.org/10.1787/db1d8e59-en

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/s15328007sem0902_5

Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational Research Methods, 18*(2), 207–230. https://doi.org/10.1177/1094428114555994

Church, A. T., & Lonner, W. J. (1998). The cross-cultural perspective in the study of personality: Rationale and current research. *Journal of Cross-Cultural Psychology*, *29*(1), 32–62. https://doi.org/10.1177/0022022198291003

Clark, L., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, *31*(12), 1412–1427. https://doi.org/10.1037/pas0000626

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis* (2nd ed. [eBook]). Psychology Press.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual*. Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (2008). The revised NEO Personality Inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment (Vol. 2): Personality measurement and testing* (pp. 179–198). Sage. https://doi.org/10.4135/9781849200479.n9

Costa, P. T., & McCrae, R. R. (2014). The NEO Inventories. In *Personality Assessment* (2nd ed.). Routledge.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.

Cronbach, L. J., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302. https://doi.org/10.1037/h0040957

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, *64*(3), 391–418. https://doi.org/10.1177/0013164404266386

Dahlsgaard, K., Peterson, C., & Seligman, M. E. P. (2005). Shared virtue: The convergence of valued human strengths across culture and history. *Review of General Psychology*, *9*(3), 203–213. https://doi.org/10.1037/1089-2680.9.3.203

Danner, D., Blasius, J., Breyer, B., Eifler, S., Menold, N., Paulhus, D., Rammstedt, D., Roberts, R., Schmitt, M., & Ziegler, M. (2016). Current challenges, new developments, and future directions in scale construction. *European Journal of Psychological Assessment*, *32*(3), 175–180. https://doi.org/10.1027/1015-5759/a000375

Danner, D., Lechner, C. M., Soto, C. J., & John, O. P. (2021). Modelling the incremental value of personality facets: The domains-incremental facets-acquiescence bifactor showmodel. *European Journal of Personality*, *35*(1), 67–84. https://doi.org/10.1002/per.2268

Danner, D., Rammstedt, B., Bluemke, M., Lechner, C. M., Berres, S., Knopf, T., Soto, C., & John, O. P. (2019). Das Big-Five Inventar 2: Validierung eines Persönlichkeitsinventars zur Erfassung von 5 Persönlichkeitsdomänen und 15 Facetten [The German Big Five Inventory 2: Measuring five personality domains and 15 facets]. *Diagnostica*, *65*(3), 121–132. https://doi.org/10.1026/0012-1924/a000218

Davis, L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research*, *5*(4), 194–197. https://doi.org/10.1016/s0897-1897(05)80008-4

De Beuckelaer, A., Weijters, B., & Rutten, A. (2010). Using ad hoc measures for response styles: A cautionary note. *Quality and Quantity*, *44*(4), 761–775. https://doi.org/10.1007/s11135-009-9225-z

Denissen, J. J. A., Luhmann, M., Chung, J. M., & Bleidorn, W. (2019). Transactions between life events and personality traits across the adult lifespan. *Journal of Personality and Social Psychology*, *116*(4), 612–633. https://doi.org/10.1037/pspp0000196

DeSimone, J. A., & Harms, P. D. (2018). Dirty data: The effects of screening respondents who provide low-quality data in survey research. *Journal of Business and Psychology*, *33*(5), 559–577. https://doi.org/10.1007/s10869-017-9514-9

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, *91*(6), 1138–1151. https://doi.org/10.1037/0022-3514.91.6.1138

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, *33*(4), 533–552. https://doi.org/10.1016/S0191-8869(01)00171-4

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*(5), 880–896. https://doi.org/10.1037/0022-3514.93.5.880

Diamond, M. R., O'Brien-Malone, A., & Woodworth, R. J. (2010). Scoring the VIA Survey of Character. *Psychological Reports*, *107*(3), 833–836. https://doi.org/10.2466/02.07.09.PR0.107.6.833-836

Diener, E. D. (1984). Subjective well-being. *Psychological Bulletin*, *95*(3), 542–575. https://doi.org/10.1037/0033-2909.95.3.542

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

du Plessis, G. A., & de Bruin, G. P. (2015). Using Rasch modelling to examine the international personality item pool (IPIP) values in action (VIA) measure of character strengths. *Journal of Psychology in Africa*, *25*(6), 512–521. https://doi.org/10.1080/14330237.2015.1124603

Duan, W., Ho, S. M. Y., Yu, B., Tang, X., Zhang, Y., Li, T., & Yuen, T. (2012). Factor structure of the Chinese virtues questionnaire. *Research on Social Work Practice*, *22*(6), 680–688. https://doi.org/10.1177/1049731512450074

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From Alpha to Omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*(6), 358–361. https://doi.org/10.1037/h0036316

Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods*, *14*(2), 370–388. https://doi.org/10.1177/1094428110378369

Eisinga, R., te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Entringer, T. M., Gebauer, J. E., & Paulhus, D. L. (2021). Extracting agency and communion from the Big Five: A four-way competition. *Assessment*. Advance online publication. https://doi.org/10.1177/10731911211003978

ESS EduNet. (n.d.). *Two Dimensions*. Retrieved October 1, 2021, from http://essedunet.nsd.uib.no/cms/topics/1/5/3.html

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fan, X., & Sivo, S. A. (2009). Using ΔGoodness-of-Fit indexes in assessing mean structure invariance. *Structural Equation Modeling*, *16*(1), 54–69. https://doi.org/10.1080/10705510802561311

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, *116*(3), 429–456. https://doi.org/10.1037/0033-2909.116.3.429

Feraco, T., Casali, N., & Meneghetti, C. (2022). Do strengths converge into virtues? An item-, virtue-, and scale-level analysis of the Italian Values in Action Inventory of Strengths-120. *Journal of Personality Assessment*, *104*(3), 395–407. https://doi.org/10.1080/00223891.2021.1934481

Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement, 78*(5), 762–780. https://doi.org/10.1177/0013164417719308

Fischer, R., & Poortinga, Y. H. (2018). Addressing methodological challenges in culture-comparative research. *Journal of Cross-Cultural Psychology*, *49*(5), 691–712. https://doi.org/10.1177/0022022117738086

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378–382. https://doi.org/10.1037/h0031619

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.2307/3151312

Fowers, B. J., Carroll, J. S., Leonhardt, N. D., & Cokelet, B. (2021). The emerging science of virtue. *Perspectives on Psychological Science*, *16*(1), 118–147. https://doi.org/10.1177/1745691620924473

Freelon, D. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science, 5*(1), 20–33.

Gander, F., Hofmann, J., & Ruch, W. (2020). Character strengths: Person–environment fit and relationships with job and life satisfaction. *Frontiers in Psychology*, *11*, Article 1582. https://doi.org/10.3389/fpsyg.2020.01582

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2011). Performance of Velicer's minimum average partial factor retention method with categorical variables. *Educational and Psychological Measurement, 71*(3), 551–570. https://doi.org/10.1177/0013164410389489

Gauckler, B., & Körner, T. (2011). Measuring the employment status in the labour force survey and the German Census 2011: Insights from recent research at Destatis.

Methoden, Daten, Analysen (Mda), *5*, 181–205. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-282136

Gennerich, C., & Huber, S. (2006). Value priorities and content of religiosity—New research perspectives. *Archiv Für Religionspsychologie/Archive for the Psychology of Religion*, *28*(1), 253–267. https://doi.org/10.1163/008467206777832508

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, *25*(2), 186–192. https://doi.org/10.2307/3172650

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, *55*(3), 377–393. https://doi.org/10.1177/0013164495055003002

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe* (Vol. 7, pp. 7–28). Tilburg University Press.

Goldberg, L. R. (2006). Doing it all Bass-Ackwards: The development of hierarchical factor structures from the top down. *Journal of Research in Personality*, *40*(4), 347–358. https://doi.org/10.1016/j.jrp.2006.01.001

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

Goldberg, L. R., & Saucier, G. (2016). *The Eugene-Springfield Community Sample: Information available from the research participants* (ORI Technical Report No. 1; Vol. 56). Oregon Research Institute.

Goodman, F. R., Disabato, D. J., & Kashdan, T. B. (2019). Integrating psychological strengths under the umbrella of personality science: Rethinking the definition, measurement, and modification of strengths. *The Journal of Positive Psychology*, *14*(1), 61–67. https://doi.org/10.1080/17439760.2018.1528380

Gorsuch, R. L. (1983–2015). *Factor analysis* (2nd ed./Classic Edition). Routledge/Taylor & Francis.

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement*, *66*(6), 930–944. https://doi.org/10.1177/0013164406288165

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, *31*(2), 197–218. https://doi.org/10.1207/s15327906mbr3102_3

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*(4), 323–343. https://doi.org/10.1037/1082-989X.11.4.323

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, *21*(4), 347–360. https://doi.org/10.1111/j.1745-3984.1984.tb01039.x

Greenleaf, E. A. (1992). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, *29*(2), 176–188. https://doi.org/10.1177/002224379202900203

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*(4), 430–450. https://doi.org/10.1037/1082-989X.6.4.430

Guion, R. (1977). Content validity: The source of my discontent. *Applied Psychological Measurement*, *1*(1), 1–10. https://doi.org/10.1177/014662167700100103

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255–282. https://doi.org/10.1007/bf02288892

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Pearson Prentice Hall.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling*, *25*(4), 621–638. https://doi.org/10.1080/10705511.2017.1402334

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sørbom (Eds.), *Structural equation modeling: Present and future—A Festschrift in honor of Karl Jöreskog* (pp. 195–216). Scientific Software International.

Harkness, J. (2003). Questionnaire translation. In J. Harkness, F. J. R. van de Vijver, & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Wiley.

Harzer, C., & Ruch, W. (2014). The role of character strengths for task performance, job dedication, interpersonal facilitation, and organizational support. *Human Performance*, *27*(3), 183–205. https://doi.org/10.1080/08959285.2014.913592

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, *9*(2), 139–164. https://doi.org/10.1177/014662168500900204

Heene, M., Bollmann, S., & Bühner, M. (2014). Much ado about nothing, or much to do about something? Effects of scale shortening on criterion validity and mean differences. *Journal of Individual Differences*, *35*(4), 245–249. https://doi.org/10.1027/1614-0001/a000146

Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, *16*(3), 319–336. https://doi.org/10.1037/a0024917

Höfer, S., Hausler, M., Huber, A., Strecker, C., Renn, D., & Höge, T. (2019). Psychometric characteristics of the German Values in Action Inventory of Strengths 120-item short form. *Applied Research in Quality of Life*, *15*, 597–611. https://doi.org/10.1007/s11482-018-9696-y

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, *14*(3), 332–346. https://doi.org/10.1177/1088868310361240

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*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., & Puranen, B. (2018). *World Values Survey: Round Six* (Country-Pooled Datafile Version:

www.worldvaluessurvey.org/WVSDocumentationWV6.jsp.). JD Systems Institute. https://doi.org/10.14281/18241.8

IPIP. (n.d.). *The Items Included in Each of the 24 Revised IPIP-VIA Scales*. Retrieved June 22, 2022, from https://ipip.ori.org/newVIAKey.htm

ISSP Research Group. (2013). *International Social Survey Programme: Work Orientations III - ISSP 2005* (ZA4350 Datenfile Version 2.0.0). GESIS Datenarchiv, Köln. https://doi.org/10.4232/1.11648

Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, *55*(4), 243–252. https://doi.org/10.1037/h0045996

Jankowsky, K., Olaru, G., & Schroeders, U. (2020). Compiling measurement invariant short scales in cross–cultural personality assessment using ant colony optimization. *European Journal of Personality*, *34*(3), 470–485. https://doi.org/10.1002/per.2260

Jennrich, R. I. (2002). A simple general method for oblique rotation. *Psychometrika*, *67*(1), 7–19. https://doi.org/10.1007/BF02294706

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger Publishers/Greenwood Publishing Group.

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality*, *2*(3), 171–203. https://doi.org/10.1002/per.2410020302

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In *Handbook of personality: Theory and research, 3rd ed* (pp. 114–158). The Guilford Press.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, *61*(4), 521–551. https://doi.org/10.1111/j.1467-6494.1993.tb00781.x

John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 461–494). Guilford.

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, *18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Kandler, C., Zimmermann, J., & McAdams, D. P. (2014). Core and surface characteristics for the description and theory of personality differences and development. *European Journal of Personality*, *28*, 231–243. https://doi.org/10.1002/per.1952

Kemper, C. J., Trapp, S., Kathmann, N., Samuel, D. B., & Ziegler, M. (2019). Short versus long scales in clinical assessment: Exploring the trade-off between resources saved and psychometric quality lost using two measures of obsessive–compulsive symptoms. *Assessment*, *26*(5), 767–782. https://doi.org/10.1177/1073191118810057

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods and Research*, *44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, *10*(3), 333–351. https://doi.org/10.1207/S15328007SEM1003_1

Klages, L. (1932). *The science of character* (p. 308). Sci-Art.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). Guilford.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lang, J. W. B., Kersting, M., & Beauducel, A. (2016). Hierarchies of factor solutions in the intelligence domain: Applying methodology from personality psychology to gain insights into the nature of intelligence. *Learning and Individual Differences*, *47*, 37–50. https://doi.org/10.1016/j.lindif.2015.12.003

Lavy, S., & Benish-Weisman, M. (2021). Character strengths as "Values in Action": Linking character strengths with values theory – an exploratory study of the case of gratitude and self-transcendence. *Frontiers in Psychology*, *12*, Article 576189. https://doi.org/10.3389/fpsyg.2021.576189

Lavy, S., & Littman-Ovadia, H. (2017). My better self: Using strengths at work and work productivity, organizational citizenship behavior, and satisfaction. *Journal of Career Development*, *44*(2), 95–109. https://doi.org/10.1177/0894845316634056

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lechner, C. M., Bhaktha, N., Groskurth, K., & Bluemke, M. (2021). Why ability point estimates can be pointless: A primer on using skill measures from large-scale assessments in secondary analyses. *Measurement Instruments for the Social Sciences*, *3*, Article 2. https://doi.org/10.1186/s42409-020-00020-5

Lechner, C. M., Partsch, M. V., Danner, D., & Rammstedt, B. (2019). Individual, situational, and cultural correlates of acquiescent responding: Towards a unified conceptual framework. *British Journal of Mathematical and Statistical Psychology*, *72*(3), 426–446. https://doi.org/10.1111/bmsp.12164

Lechner, C. M., & Rammstedt, B. (2015). Cognitive ability, acquiescence, and the structure of personality in a sample of older adults. *Psychological Assessment*, *27*(4), 1301–1311. https://doi.org/10.1037/pas0000151

Lee, K., & Ashton, M. C. (2006). Further assessment of the HEXACO Personality Inventory: Two new facet scales and an observer report form. *Psychological Assessment*, *18*(2), 182–191. https://doi.org/10.1037/1040-3590.18.2.182

Lee, K., & Ashton, M. C. (2008). The HEXACO personality factors in the indigenous personality lexicons of English and 11 other languages. *Journal of Personality*, *76*(5), 1001–1054. https://doi.org/10.1111/j.1467-6494.2008.00512.x

Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment*, *25*(5), 543–556. https://doi.org/10.1177/1073191116659134

Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods*, *13*(3), 229–248. https://doi.org/10.18148/srm/2019.v13i3.7403

Lim, S., & Jahng, S. (2019). Determining the number of factors using parallel analysis and its recent variants. *Psychological Methods*, *24*(4), 452–467. https://doi.org/10.1037/met0000230

Lins De Holanda Coelho, G., Hanel, P. H. P., Johansen, M. K., & Maio, G. R. (2021). Mental representations of values and behaviors. *European Journal of Personality*. Advance online publication. https://doi.org/10.1177/08902070211034385

Littman-Ovadia, H. (2015). Brief report: Short form of the VIA Inventory of Strengths: Construction and initial tests of reliability and validity. *International Journal of Humanities Social Sciences and Education*, *2*(4), 229–237.

Littman-Ovadia, H., & Lavy, S. (2012). Character strengths in Israel: Hebrew adaptation of the VIA Inventory of Strengths. *European Journal of Psychological Assessment*, *28*(1), 41–50. https://doi.org/10.1027/1015-5759/a000089

Liu, D. (2012). Translation and culture: Translating idioms between English and Chinese from a cultural perspective. *Theory and Practice in Language Studies*, *2*(11), 2357–2362. https://doi.org/10.4304/tpls.2.11.2357-2362

Lorenzo-Seva, U. (2013). *How to report the percentage of explained common variance in exploratory factor analysis* [Technical Report]. Department of Psychology, Universitat Rovira i Virgili. http://psico.fcep.urv.es/utilitats/factor/documentation/Percentage_of_explained_common_variance.pdf

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavioral Research Methods*, *38*, 88–91. https://doi.org/10.3758/BF03192753

Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2—A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT models. *Applied Psychological Measurement*, *37*(6), 497–498. https://doi.org/10.1177/0146621613487794

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57–64. https://doi.org/10.1027/1614-2241.2.2.57

Macdonald, C., Bore, M., & Munro, D. (2008). Values in action scale and the Big 5: An empirical indication of structure. *Journal of Research in Personality*, *42*(4), 787–799. https://doi.org/10.1016/j.jrp.2007.10.003

MacDonald, D. A., Friedman, H. L., Brewczynski, J., Holland, D., Salagame, K. K. K., Mohan, K. K., Gubrij, Z. O., & Cheong, H. W. (2015). Spirituality as a scientific construct: Testing its universality across cultures and languages. *PLoS ONE*, *10*(3), Article e0117701. https://doi.org/10.1371/journal.pone.0117701

Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

Martínez-Martí, M. L., & Ruch, W. (2017). Character strengths predict resilience over and above positive affect, self-efficacy, optimism, social support, self-esteem, and life satisfaction. *Journal of Positive Psychology*, *12*(2), 110–119. https://doi.org/10.1080/17439760.2016.1163403

McCrae, R. R., & Costa, P. T. (2008). The five-factor theory of personality. In *Handbook of personality: Theory and research* (3rd ed., pp. 159–181). The Guilford Press.

McCrae, R. R., Herbst, J. H., & Costa, P. T., Jr. (2001). Effects of acquiescence on personality factor structures. In R. Riemann, F. Ostendorf, & F. Spinath (Eds.), *Personality and temperament: Genetics, evolution, and structure* (pp. 217–231). Pabst Science.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, *15*(1), 28–50. https://doi.org/10.1177/1088868310366253

McCrae, R. R., & Mõttus, R. (2019). What personality scales measure: A new psychometrics and its implications for theory and assessment. *Current Directions in Psychological Science*, *28*(4), 415–420. https://doi.org/10.1177/0963721419849559

McCullough, M. E., Emmons, R. A., & Tsang, J.-A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology*, *82*(1), 112–127. https://doi.org/10.1037/0022-3514.82.1.112

McDonald, R. P. (1999). *Test theory: A unified treatment.* Erlbaum.

McGrath, R. E. (2014). Scale- and item-level factor analyses of the VIA Inventory of Strengths. *Assessment*, *21*(1), 4–14. https://doi.org/10.1177/1073191112450612

McGrath, R. E. (2015). Integrating psychological and cultural perspectives on virtue: The hierarchical structure of character strengths. *The Journal of Positive Psychology*, *10*(5), 407–424. https://doi.org/10.1080/17439760.2014.994222

McGrath, R. E. (2016). Measurement invariance in translations of the VIA inventory of strengths. *European Journal of Psychological Assessment*, *32*(3), 187–194. https://doi.org/10.1027/1015-5759/a000248

McGrath, R. E. (2019). *Technical report: The VIA Assessment Suite for Adults: Development and initial evaluation* (rev ed.). VIA Institute on Character. https://www.viacharacter.org/pdf/Technical%20Report%20Revised%20Edition%202019_1.pdf

McGrath, R. E. (2022). A summary of construct validity evidence for two measures of character strengths. *Journal of Personality Assessment*. Advance online publication. https://doi.org/10.1080/00223891.2022.2120402

McGrath, R. E., Brown, M., Westrich, B., & Han, H. (2022). Representative sampling of the VIA Assessment Suite for Adults. *Journal of Personality Assessment*, *104*(3), 380–394. https://doi.org/10.1080/00223891.2021.1955692

McGrath, R. E., Greenberg, M. J., & Hall-Simmonds, A. (2018). Scarecrow, Tin Woodsman, and Cowardly Lion: The three-factor model of virtue. *The Journal of Positive Psychology*, *13*(4), 373–392. https://doi.org/10.1080/17439760.2017.1326518

McGrath, R. E., Hall-Simmonds, A., & Goldberg, L. R. (2020). Are measures of character and personality distinct? Evidence from observed-score and true-score analyses. *Assessment*, *27*(1), 117–135. https://doi.org/10.1177/1073191117738047

McGrath, R. E., & Wallace, N. (2021). Cross-validation of the VIA Inventory of Strengths-Revised and its short forms. *Journal of Personality Assessment*, *103*(1), 120–131. https://doi.org/10.1080/00223891.2019.1705465

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/bf02294825

Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Toward modern psychometrics: Application of item response theory models in personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality* (pp. 407–423). Guilford.

Moshagen, M., & Auerswald, M. (2018). On congruence and incongruence of measures of fit in structural equation modeling. *Psychological Methods*, *23*(2), 318–336. https://doi.org/10.1037/met0000122

Mõttus, R., & Allerhand, M. H. (2018). Why do traits come together? The underlying trait and network approaches. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *The SAGE handbook of personality and individual differences: The science of personality and individual differences* (pp. 130–151). SAGE Publications Ltd. https://doi.org/10.4135/9781526451163.n6

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *112*(3), 474–490. https://doi.org/10.1037/pspp0000100

Mõttus, R., Wood, D., Condon, D. M., Back, M. D., Baumert, A., Costantini, G., Epskamp, S., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Yarkoni, T., Ziegler, M., & Zimmermann, J. (2020). Descriptive, predictive and explanatory personality research: Different goals, different approaches, but a shared need to move beyond the Big Few traits. *European Journal of Personality*, *34*(6), 1175–1201. https://doi.org/10.1002/per.2311

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes* [Unpublished technical report]. College of Education, UCLA. https://www.statmodel.com/download/Article_075.pdf

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus User's Guide* (version 7, rev.3). Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus User's Guide* (8th ed.). Muthén & Muthén.

Najderska, M., & Cieciuch, J. (2018). The structure of character strengths: Variable- and person-centered approaches. *Frontiers in Psychology*, *9*, Article 153. https://doi.org/10.3389/fpsyg.2018.00153

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. The National Academies Press. https://doi.org/10.17226/13398

Nesselroade, J. R., & McArdle, J. J. (1997). On the mismatching of levels of abstraction in mathematical-statistical model fitting. In H. W. Reese & M. D. Franzen (Eds.), *Biological and neuropsychological mechanisms: Life-span developmental psychology* (pp. 23–49). Lawrence Erlbaum Associates, Inc.

Ng, V., Cao, M., Marsh, H. W., Tay, L., & Seligman, M. E. P. (2017). The factor structure of the Values in Action Inventory of Strengths (VIA-IS): An item-level exploratory structural equation modeling (ESEM) bifactor analysis. *Psychological Assessment*, *29*(8), 1053–1058. https://doi.org/10.1037/pas0000396

Nießen, D., Groskurth, K., Rammstedt, B., & Lechner, C. M. (2020). *An English-language adaptation of the General Life Satisfaction Short Scale (L-1)*. Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/zis284

Noftle, E. E., Schnitker, S. A., & Robins, R. W. (2011). Character and personality: Connections between positive psychology and personality psychology. In K. M. Sheldon, T. B. Kashdan, & M. F. Steger (Eds.), *Designing positive psychology: Taking stock and moving forward* (pp. 207–227). Oxford University Press.

Norman, W. T. (1967). *2800 personality trait descriptors: Normative operating characteristics for a university population*. Department of Psychology, University of Michigan.

Noronha, A. P. P., Dellazzana-Zanon, L. L., & Zanon, C. (2015). Internal structure of the characters strengths scale in Brazil. *Psico-USF*, *20*(2), 229–235. https://doi.org/10.1590/1413-82712015200204

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(3), 396–402. https://doi.org/10.3758/BF03200807

O'Connor, B. P. (2020). *EFA.dimensions: Exploratory factor analysis functions for assessing dimensionality* (Version 0.1.6.) [R package]. https://CRAN.R-project.org/package=EFA.dimensions

Olaru, G., & Danner, D. (2021). Developing cross-cultural short scales using ant colony optimization. *Assessment*, *28*(1), 199–210. https://doi.org/10.1177/1073191120918026

Olaru, G., & Jankowsky, K. (2021). The HEX-ACO-18: Developing an age-invariant HEXACO short scale using ant colony optimization. *Journal of Personality*

Assessment. Advance online publication. https://doi.org/10.1080/00223891.2021.1934480

Olaru, G., Schroeders, U., Hartung, J., & Wilhelm, O. (2019). Ant colony optimization and local weighted structural equation modeling. A tutorial on novel item and person sampling procedures for personality research. *European Journal of Personality*, *33*(3), 400–419. https://doi.org/10.1002/per.2195

Osborne, J., & Costello, A. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research, and Evaluation*, *9*, Article 11. https://doi.org/10.7275/ktzq-jq66

Park, N. (2018). Classifying and measuring strengths of character. In C. R. Snyder, S. J. Lopez, L. M. Edwards, & S. C. Marques (Eds.), *Oxford handbook of positive psychology* (3rd ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199396511.001.0001

Park, N., & Peterson, C. (2010). Does it matter where we live? The urban psychology of character strengths. *American Psychologist*, *65*(6), 535–547. https://doi.org/10.1037/a0019621

Park, N., Peterson, C., & Seligman, M. E. P. (2004). Strengths of character and well-being. *Journal of Social and Clinical Psychology*, *23*(5), 603–619. https://doi.org/10.1521/jscp.23.5.603.50748

Park, N., Peterson, C., & Seligman, M. E. P. (2006). Character strengths in fifty-four nations and the fifty US states. *Journal of Positive Psychology*, *1*(3), 118–129. https://doi.org/10.1080/17439760600619567

Partsch, M. V., Bluemke, M., & Lechner, C. M. (2022). Revisiting the hierarchical structure of the 24 VIA character strengths: Three global dimensions may suffice to capture their essence. *European Journal of Personality*, *36*(5), 825–845. https://doi.org/10.1177/08902070211017760

Peterson, C. (2003). *The Values-in-Action Structured Interview (VIA-SI)*. The VIA Institute on Character.

Peterson, C., & Park, N. (2009). Classifying and measuring strengths of character. In S. J. Lopez & C. R. Snyder (Eds.), *Oxford handbook of positive psychology* (2nd ed., pp. 25–33). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195187243.001.0001

Peterson, C., Park, N., Pole, N., D'Andrea, W., & Seligman, M. E. P. (2008). Strengths of character and posttraumatic growth. *Journal of Traumatic Stress*, *21*(2), 214–217. https://doi.org/10.1002/jts.20332

Peterson, C., Park, N., & Seligman, M. E. P. (2005). Assessment of character strengths. In G. P. Koocher, J. C. Norcross, & S. S. Hill III (Eds.), *Psychologists' desk reference* (2nd ed., pp. 93–98). Oxford University Press.

Peterson, C., & Seligman, M. E. P. (2003). Character strengths before and after September 11. *Psychological Science*, *14*(4), 381–384. https://doi.org/10.1111/1467-9280.24482

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. American Psychological Association.

Petkari, E., & Ortiz-Tallo, M. (2018). Towards youth happiness and mental health in the United Arab Emirates: The path of character strengths in a multicultural population. *Journal of Happiness Studies*, *19*(2), 333–350. https://doi.org/10.1007/s10902-016-9820-3

Piedmont, R. L. (2007). Cross-cultural generalizability of the Spiritual Transcendence Scale to the Philippines: Spirituality as a human universal. *Mental Health, Religion and Culture*, *10*(2), 89–107. https://doi.org/10.1080/13694670500275494

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*(1), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing and Health*, *30*(4), 459–467. https://doi.org/10.1002/nur.20199

Pozzebon, J. A., Visser, B. A., Ashton, M. C., Lee, K., & Goldberg, L. R. (2010). Psychometric characteristics of a public-domain self-report measure of vocational interests: The Oregon Vocational Interest Scales. *Journal of Personality Assessment*, *92*(2), 168–174. https://doi.org/10.1080/00223890903510431

Primi, R., De Fruyt, F., Santos, D., Antonoplis, S., & John, O. P. (2020). True or false? Keying direction and acquiescence influence the validity of socio-emotional skills items in predicting high school achievement. *International Journal of Testing*, *20*(2), 97–121. https://doi.org/10.1080/15305058.2019.1673398

Proyer, R. T., Gander, F., Wellenzohn, S., & Ruch, W. (2013). What good are character strengths beyond subjective well-being? The contribution of the good character on self-reported health-oriented behavior, physical fitness, and the subjective health status. *The Journal of Positive Psychology*, *8*(3), 222–232. https://doi.org/10.1080/17439760.2013.777767

Proyer, R. T., Gander, F., Wyss, T., & Ruch, W. (2011). The relation of character strengths to past, present, and future life satisfaction among German-speaking women. *Applied Psychology: Health and Well-Being*, *3*(3), 370–384. https://doi.org/10.1111/j.1758-0854.2011.01060.x

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team. (2018). *R: A language and environment for statistical computing* (Version 3.4.4) [Statistical software]. R Foundation for Statistical Computing. http://www.R-project.org/

R Core Team. (2019). *R: A language and environment for statistical computing* (Version 3.6.1) [Statistical software]. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2020). *R: A language and environment for statistical computing* (Version 4.0.2) [Statistical software]. R Foundation for Statistical Computing. https://www.R-project.org/

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. https://doi.org/10.2307/271063

Rammstedt, B., & Beierlein, C. (2014). Can't we make it any shorter? The limits of personality assessment and ways to overcome them. *Journal of Individual Differences*, *35*(4), 212–220. https://doi.org/10.1027/1614-0001/a000141

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2020). Validation of the short and extra-short forms of the Big Five Inventory-2 (BFI-2) and their German adaptations. *European Journal of Psychological Assessment*, *36*(1), 149–161. https://doi.org/10.1027/1015-5759/a000481

Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment*, *25*(4), 1137–1145. https://doi.org/10.1037/a0033323

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Rammstedt, B., Kemper, C. J., & Borg, I. (2013). Correcting Big Five personality measurements for acquiescence: An 18–country cross–cultural study. *European Journal of Personality*, *27*(1), 71–81. https://doi.org/10.1002/per.1894

Rammstedt, B., Lechner, C. M., & Danner, D. (2021). Short forms do not fall short: A comparison of three (extra-)short forms of the Big Five. *European Journal of Psychological Assessment*, *37*(1), 23–32. https://doi.org/10.1027/1015-5759/a000574

Rashid, T. (2015). Positive psychotherapy: A strength-based approach. *The Journal of Positive Psychology*, *10*(1), 25–40. https://doi.org/10.1080/17439760.2014.920411

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*(2), 173–184. https://doi.org/10.1177/01466216970212006

Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement*, *22*(4), 375–385. https://doi.org/10.1177/014662169802200407

Raykov, T. (2001). Bias of coefficient alpha for fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, *25*(1), 69–76. https://doi.org/10.1177/01466216010251005

Redfern, K., Simpson, A., & Ran, Z. (2014, December 3–5). *Character strengths in employees in the People's Republic of China: Analysing the factor structure of the VIA Inventory of Strengths* [Paper presentation]. 28th Australian and New Zealand Academy of Management Conference: Reshaping Management for Impact, UTS Business School, Sydney, Australia.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Respondi. (2022, June 22). *Access Panels & Services for Digital Studies*. https://www.respondi.com/EN/

Revelle, W. (2019a). *Fa.parallel function* [R Documentation]. https://www.rdocumentation.org/packages/psych/versions/1.9.12.31/topics/fa.parallel

Revelle, W. (2019b). *psych: Procedures for psychological, psychometric, and personality research* (Version 1.8.12) [R package]. Northwestern University. http://CRAN.R-project.org/package=psych

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rose, N., Wagner, W., Mayer, A., & Nagengast, B. (2019). Model-based manifest and latent composite scores in structural equation models. *Collabra: Psychology*, *5*(1), Article 9. https://doi.org/10.1525/collabra.143

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

Ruch, W., Gander, F., Wagner, L., & Giuliani, F. (2019). The structure of character: On the relationships between character strengths and virtues. *The Journal of Positive Psychology*. https://doi.org/10.1080/17439760.2019.1689418

Ruch, W., & Heintz, S. (2016). The virtue gap in humor: Exploring benevolent and corrective humor. *Translational Issues in Psychological Science*, *2*(1), 35–45. https://doi.org/10.1037/tps0000063

Ruch, W., Huber, A., Beermann, U., & Proyer, R. T. (2007). Character strengths as predictors of the “good life” in Austria, Germany and Switzerland. In Romanian Academy - “George Barit” Department of Social Research (Ed.), *Studies and Researches in Social Sciences* (pp. 123–131). Academia Română. https://doi.org/10.5167/uzh-3648

Ruch, W., Martínez-Martí, M. L., Proyer, R. T., & Harzer, C. (2014). The Character Strengths Rating Form (CSRF): Development and initial assessment of a 24-Item rating scale to assess character strengths. *Personality and Individual Differences*, *68*, 53–58. https://doi.org/10.1016/j.paid.2014.03.042

Ruch, W., & Proyer, R. T. (2015). Mapping strengths into virtues: The relation of the 24 VIA-strengths to six ubiquitous virtues. *Frontiers in Psychology*, *6*, Article 460. https://doi.org/10.3389/fpsyg.2015.00460

Ruch, W., Proyer, R. T., Harzer, C., Park, N., Peterson, C., & Seligman, M. E. P. (2010). Values in Action Inventory of Strengths (VIA-IS): Adaptation and validation of the German version and the development of a peer-rating form. *Journal of Individual Differences*, *31*(3), 138–149. https://doi.org/10.1027/1614-0001/a000022

Ruch, W., & Stahlmann, A. G. (2020). 15 years after Peterson and Seligman (2004): A brief narrative review of the research on the 12 criteria for character strengths–the forgotten treasure of the VIA classification. In M. Brohm-Badry, C. Peifer, J. M. Greve, & B. Berend (Eds.), *Zusammenwachsen – Förderung der positiv-psychologischen Entwicklung von Individuen,Organisationen und Gesellschaft* (pp. 142–172). Pabst Science Publishers.

Ruch, W., Vylobkova, V., & Heintz, S. (2021). *Two of a kind or distant relatives? A multimethod investigation of the overlap between personality traits and character strengths*. PsyArXiv. https://doi.org/10.31234/osf.io/93pbk

Rudnev, M., Magun, V., & Schmidt, P. (2016). Basic human values: Stability of value typology in Europe. In M. Voicu, I. C. Mochmann, & H. Dülmer (Eds.), *Values, Economic Crisis and Democracy* (1st ed., pp. 21–49). Routledge.

Rushton, J. P., Bons, T. A., & Hur, Y.-M. (2008). The genetics and evolution of the general factor of personality. *Journal of Research in Personality*, *42*(5), 1173–1185. https://doi.org/10.1016/j.jrp.2008.03.002

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement, 74*(1), 31–57. https://doi.org/10.1177/0013164413498257

Saroglou, V. (2011). Believing, Bonding, Behaving, and Belonging: The big four religious dimensions and cultural variation. *Journal of Cross-Cultural Psychology*, *42*(8), 1320–1340. https://doi.org/10.1177/0022022111412267

Saroglou, V., Delpierre, V., & Dernelle, R. (2004). Values and religiosity: A meta-analysis of studies using Schwartz's model. *Personality and Individual Differences*, *37*(4), 721–734. https://doi.org/10.1016/j.paid.2003.10.005

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*, 243–248. https://doi.org/10.1007/s11336-009-9135-y

Saucier, G. (2009). Recurrent personality dimensions in inclusive lexical studies: Indications for a Big Six structure. *Journal of Personality*, *77*(5), 1577–1614. https://doi.org/10.1111/j.1467-6494.2009.00593.x

Saucier, G. (2019). On the curious history of the Big Six. *Zeitschrift Für Psychologie*, *227*(3), 166–173. https://doi.org/10.1027/2151-2604/a000375

Saucier, G., & Goldberg, L. R. (1996). The language of personality: Lexical perspectives on the five-factor model. In J. S. Wiggins (Ed.), *The Five-Factor Model of personality* (pp. 21–50). Guilford.

Saucier, G., & Goldberg, L. R. (1998). What is beyond the Big Five? *Journal of Personality*, *66*(4), 495–524. https://doi.org/10.1111/1467-6494.00022

Saucier, G., & Iurino, K. (2020). High-dimensionality personality structure in the natural language: Further analysis of classic sets of English-language trait-adjectives. *Journal of Personality and Social Psychology*, *119*(5), 1188–1219. https://doi.org/10.1037/pspp0000273

Saucier, G., & Skrzypińska, K. (2006). Spiritual but not religious? Evidence for two independent dispositions. *Journal of Personality*, *74*(5), 1257–1292. https://doi.org/10.1111/j.1467-6494.2006.00409.x

Saucier, G., Thalmayer, A. G., Payne, D. L., Carlson, R., Sanogo, L., Ole-Kotikash, L., Church, A. T., Katigbak, M. S., Somer, O., Szarota, P., Szirmák, Z., & Zhou, X. (2014). A basic bivariate structure of personality attributes evident across nine languages. *Journal of Personality*, *82*(1), 1–14. https://doi.org/10.1111/jopy.12028

Savalei, V. (2018). On the computation of the RMSEA and CFI from the mean-and-variance corrected test statistic with nonnormal data in SEM. *Multivariate Behavioral Research*, *53*(3), 419–429. https://doi.org/10.1080/00273171.2018.1455142

Schimmack, U., Böckenholt, U., & Reisenzein, R. (2002). Response styles in affect ratings: Making a mountain out of a molehill. *Journal of Personality Assessment*, *78*(3), 461–483. https://doi.org/10.1207/S15327752JPA7803_06

Schmidt, P., Bamberg, S., Davidov, E., Herrmann, J., & Schwartz, S. H. (2007). Die Messung von Werten mit dem "Portraits Value Questionnaire" [The measurement of values with the "Portrait Value Questionnaire"]. *Zeitschrift Für Sozialpsychologie*, *38*(4), 261–275. https://doi.org/10.1024/0044-3514.38.4.261

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Schroeders, U., Wilhelm, O., & Olaru, G. (2016). Meta-heuristics in short scale construction: Ant colony optimization and genetic algorithm. *PLoS ONE*, *11*(11), Article e0167110. https://doi.org/10.1371/journal.pone.0167110

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Lawrence Erlbaum Associates.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1–65). Academic Press. https://doi.org/10.1016/S0065-2601(08)60281-6

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues*, *50*(4), 19–45. https://doi.org/10.1111/j.1540-4560.1994.tb01196.x

Schwartz, S. H. (2003a). A proposal for measuring value orientations across nations. In *Questionnaire development package of the European Social Survey* (pp. 259–319). https://www.europeansocialsurvey.org/docs/methodology/core_ess_questionnaire/ESS_core_questionnaire_human_values.pdf

Schwartz, S. H. (2003b). *Computing scores for the 10 human values*. European Social Survey. https://www.europeansocialsurvey.org/docs/methodology/ESS_computing_human_values_scale.pdf

Schwartz, S. H. (2006). A theory of cultural value orientations: Explication and applications. *Comparative Sociology*, *5*(2–3), 137–182. https://doi.org/10.1163/156913306778667357

Schwartz, S. H. (2012). An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture*, *2*(1), Article 11. https://doi.org/10.9707/2307-0919.1116

Schwartz, S. H. (2013). *Human Values*. https://essedunet.nsd.uib.no/cms/topics/1/

Schwartz, S. H. (2015). Basic individual values: Sources and consequences. In T. Brosch & D. Sander (Eds.), *Handbook of value: Perspectives from economics, neuroscience,* *philosophy, psychology and sociology* (1st ed., pp. 63–84). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198716600.003.0004

Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology*, *53*(3), 550–562. https://doi.org/10.1037/0022-3514.53.3.550

Schwartz, S. H., & Bilsky, W. (1990). Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology*, *58*(8), 878–891. https://doi.org/10.1037/0022-3514.58.5.878

Schwartz, S. H., & Boehnke, K. (2004). Evaluating the structure of human values with confirmatory factor analysis. *Journal of Research in Personality*, *38*(3), 230–255. https://doi.org/10.1016/s0092-6566(03)00069-2

Schwartz, S. H., Breyer, B., & Danner, D. (2015). *Human Values Scale (ESS)*. Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/zis234

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J. E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, *103*(4), 663–688. https://doi.org/10.1037/a0029393

Schwartz, S. H., & Huismans, S. (1995). Value priorities and religiosity in four Western religions. *Social Psychology Quarterly*, *58*(2), 88–107. https://doi.org/10.2307/2787148

Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., & Harris, M. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross Cultural Psychology*, *32*(5), 519–542. https://doi.org/10.1177/0022022101032005001

Schwartz, S. H., & Sagiv, L. (1995). Identifying culture-specifics in the content and structure of values. *Journal of Cross-Cultural Psychology*, *26*(1), 92–116. https://doi.org/10.1177/0022022195261007

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Seibel, B. L., DeSousa, D., & Koller, S. H. (2015). Adaptação Brasileira e Estrutura Fatorial da Escala 240-item VIA Inventory of Strengths [Brazilian adaptation and factor structure of the 240-item VIA Inventory of Strengths scale]. *Psico-USF*, *20*(3), 371–383. https://doi.org/10.1590/1413-82712015200301

Seligman, M. E. P. (2002). *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. Free Press/Simon and Schuster.

Seligman, M. E. P. (2011). *Flourish: A visionary new understanding of happiness and well-being*. Free Press/Simon and Schuster.

Seligman, M. E. P., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. *American Psychologist*, *60*(5), 410–421. https://doi.org/10.1037/0003-066X.60.5.410

Shapiro, A., & Ten Berge, J. M. (2002). Statistical inference of minimum rank factor analysis. *Psychometrika*, *67*(1), 79–94. https://doi.org/10.1007/BF02294710

Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement*, *79*(2), 310–334. https://doi.org/10.1177/0013164418783530

Shryack, J., Steger, M. F., Krueger, R. F., & Kallie, C. S. (2010). The structure of virtue: An empirical investigation of the dimensionality of the virtues in action inventory of strengths. *Personality and Individual Differences*, *48*(6), 714–719. https://doi.org/10.1016/j.paid.2010.01.007

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Singh, K., & Choubisa, R. (2009). Psychometric properties of Hindi translated version of Values in Action Inventory of Strengths (VIA-IS). *Journal of Indian Health Psychology*, *4*(1), 65–76.

Singh, K., & Choubisa, R. (2010). Empirical validation of values in action-inventory of strengths (VIA-IS) in Indian context. *Psychological Studies*, *55*(2), 151–158. https://doi.org/10.1007/s12646-010-0015-4

Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research*, *45*(1/3), 83–117. https://doi.org/10.1023/A:1006985528729

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*(1), 102–111. https://doi.org/10.1037//1040-3590.12.1.102

Soto, C. J., & John, O. P. (2017a). 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

Soto, C. J., & John, O. P. (2017b). Short and extra-short forms of the Big Five Inventory-2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, *68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Soto, C. J., & John, O. P. (2019). Optimizing the length, width, and balance of a personality scale: How do internal characteristics affect external validity? *Psychological Assessment*, *31*(4), 444–459. https://doi.org/10.1037/pas0000586

Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of Personality and Social Psychology, 101*(4), 862–882. https://doi.org/10.1037/a0024950

Stahlmann, A. G., & Ruch, W. (2020). Scrutinizing the criteria for character strengths: Laypersons assert that every strength is positively morally valued, even in the absence of tangible outcomes. *Frontiers in Psychology*, *11*, Article 591028. https://doi.org/10.3389/fpsyg.2020.591028

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4

Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *9*(1), 1–12. https://doi.org/10.1027/1614-2241/a000049

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, *52*(4), 589–617. https://doi.org/10.1007/bf02294821

Strus, W., & Cieciuch, J. (2017). Towards a synthesis of personality, temperament, motivation, emotion and mental health models within the Circumplex of Personality Metatraits. *Journal of Research in Personality*, *66*, 70–95. https://doi.org/10.1016/j.jrp.2016.12.002

Strus, W., & Cieciuch, J. (2019). Are the questionnaire and the psycho-lexical Big Twos the same? Towards an integration of personality structure within the Circumplex of Personality Metatraits. *International Journal of Personality Psychology*, *5*, 18–35. https://doi.org/10.21827/ijpp.5.35594

Strus, W., & Cieciuch, J. (2021). The Circumplex of Personality Metatraits and the HEXACO model: Toward refinement and integration. *Journal of Personality*, *89*(4), 803–818. https://doi.org/10.1111/jopy.12616

Strus, W., Cieciuch, J., & Rowiński, T. (2014). The Circumplex of Personality Metatraits: A synthesizing model of personality based on the Big Five. *Review of General Psychology, 18*(4), 273–286. https://doi.org/10.1037/gpr0000017

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Ten Berge, J. M. F. (1998). Some recent developments in factor analysis and the search for proper communalities. In A. Rizzi, M. Vichi, & H.-H. Bock (Eds.), *Advances in data science and classification* (pp. 325–334). Springer.

Ten Berge, J. M. F., & Kiers, H. A. L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika, 56*(2), 309–315.

Thalmayer, A. G., & Saucier, G. (2014). The questionnaire Big Six in 26 nations: Developing cross–culturally applicable Big Six, Big Five and Big Two Inventories. *European Journal of Personality*, *28*(5), 482–496. https://doi.org/10.1002/per.1969

Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of brief to medium-length Big Five and Big Six Personality Questionnaires. *Psychological Assessment*, *23*(4), 995–1009. https://doi.org/10.1037/a0024165

Tilden, V., Nelson, C., & May, B. (1990). Use of qualitative methods to enhance content validity. *Nursing Research*, *39*(3), 172–175. https://doi.org/10.1097/00006199-199005000-00015

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of order polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209–220. https://doi.org/10.1037/a0023353

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (Personnel Research Section Report No. 984). Department of the Army.

Ushey, K. (2020). *renv: Project Environments* (Version 0.12.0) [R package]. https://CRAN.R-project.org/package=renv

Valentini, F. (2017). Influência e controle da aquiescência na análise fatorial. *Revista Avaliação Psicológica*, *16*(2), 116–119. https://doi.org/10.15689/ap.2017.1602.ed

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), 4–69. https://doi.org/10.1177/109442810031002

Vanhove, A. J., Harms, P. D., & DeSimone, J. A. (2016). The Abbreviated Character Strengths Test (ACST): A preliminary assessment of test validity. *Journal of Personality Assessment*, *98*(5), 536–544. https://doi.org/10.1080/00223891.2016.1148044

Vazire, S. (2010). Who knows what about a person? The Self-Other Knowledge Asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*(2), 281–300. https://doi.org/10.1037/a0017908

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. https://doi.org/10.1007/BF02293557

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). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4615-4397-8_3

VIA Institute on Character. (n.d.). *VIA Assessments*. Retrieved June 22, 2020, from https://www.viacharacter.org/researchers/assessments

Vylobkova, V., Heintz, S., Gander, F., Wagner, L., & Ruch, W. (2023). Convergence and psychometric properties of character strengths measures: The VIA-IS and the VIA-IS-

R. *Journal of Personality Assessment*, *105*(1), 14–21. https://doi.org/10.1080/00223891.2022.2044342

Wagner, L., Holenstein, M., Wepf, H., & Ruch, W. (2020). Character strengths are related to students' achievement, flow experiences, and enjoyment in teacher-centered learning, individual, and group work beyond cognitive ability. *Frontiers in Psychology*, *11*, Article 1324. https://doi.org/10.3389/fpsyg.2020.01324

Wagner, L., & Ruch, W. (2015). Good character at school: Positive classroom behavior mediates the link between character strengths and school achievement. *Frontiers in Psychology*, *6*, Article 610. https://doi.org/10.3389/fpsyg.2015.00610

Waller, N. (2007). A general method for computing hierarchical component structures by Goldberg's Bass-Ackwards method. *Journal of Research in Personality*, *41*(4), 745–752. https://doi.org/10.1016/j.jrp.2006.08.005

Wänke, M. (2002). Conversational norms and the interpretation of vague quantifiers. *Applied Cognitive Psychology*, *16*(3), 301–307. https://doi.org/10.1002/acp.787

Watkins, P. C. (2016). *Positive Psychology 101*. Springer. https://doi.org/10.1891/9780826126986

Weber, M., Ruch, W., Littman-Ovadia, H., Lavy, S., & Gai, O. (2013). Relationships among higher-order strengths factors, subjective well-being, and general self-efficacy – The case of Israeli adolescents. *Personality and Individual Differences*, *55*(3), 322–327. https://doi.org/10.1016/j.paid.2013.03.006

Weber, M., Wagner, L., & Ruch, W. (2016). Positive feelings at school: On the relationships between students' character strengths, school-related affect, and school functioning. *Journal of Happiness Studies*, *17*(1), 341–355. https://doi.org/10.1007/s10902-014-9597-1

Webster's new international dictionary of the English language. (1928). Merriam Webster.

Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological Methods*, *15*(1), 96–110. https://doi.org/10.1037/a0018721

Widaman, K. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, *28*(3), 263–311. https://doi.org/10.1207/s15327906mbr2803_1

Wittmann, W. W. (1988). Multivariate reliability theory. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 505–560). Springer US. https://doi.org/10.1007/978-1-4613-0893-5_16

Wright, J. C., Warren, M. T., & Snow, N. E. (2020). *Understanding virtue: Theory and measurement*. Oxford University Press.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multigroup confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, *12*(3), 1–26. https://doi.org/10.7275/mhqa-cd89

WVS. (n.d.). *World Values Survey*. Retrieved June 22, 2022, from https://www.worldvaluessurvey.org/wvs.jsp

Xie, W. (2015). *An investigation of virtues (derived from character strengths) in relation to psychological adjustment among Chinese international college students* [Doctoral dissertation, University of Minnesota]. University of Minnesota Digital Conservancy. http://hdl.handle.net/11299/175375

Ziegler, M. (2014). Stop and state your intentions! Let’s not forget the ABC of test construction. *European Journal of Psychological Assessment*, *30*(4), 239–242. https://doi.org/10.1027/1015-5759/a000228

Ziegler, M., & Bensch, D. (2013). Lost in translation: Thoughts regarding the translation of existing psychological measures into other languages. *European Journal of Psychological Assessment*, *29*(2), 81–83. https://doi.org/10.1027/1015-5759/a000167

Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short scales – Five misunderstandings and ways to overcome them. *Journal of Individual Differences*, *35*(4), 185–189. https://doi.org/10.1027/1614-0001/a000148

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432–442. https://doi.org/10.1037/0033-2909.99.3.432

STATEMENT OF ORIGINALITY

I hereby declare that I completed this doctoral thesis independently based on the stated resources and aids. I have not applied for a doctoral degree elsewhere and do not have a corresponding doctoral degree. I have not submitted this doctoral thesis, or parts of it, to another academic institution and this thesis has not been accepted or rejected.

Studies 1–3 of this dissertation are either published or intended for publication. Author contributions are presented below.

PUBLICATION STATUS	AUTHOR CONTRIBUTIONS (Based on the CRediT Classification)
A previous version of **Study 1** was informally published as Bluemke, M., Partsch, M. V., Saucier, G., & Lechner, C. M. (2021). *Human character in the IPIP: Towards shorter, more content-valid, and cross-culturally comparable IPIP-VIA character strength scales*. PsyArXiv. https://doi.org/10.31234/osf.io/k79qf Note: I made several minor revisions to this version for inclusion in this dissertation.	**Matthias Bluemke:** Conceptualization (equal), data curation (equal), formal analysis (lead), investigation (equal), methodology (equal), validation (equal), visualization (lead), writing—original draft (equal). **Melanie V. Partsch:** Conceptualization (equal), data curation (equal), formal analysis (equal), investigation (lead), methodology (equal), validation (equal), writing—original draft (equal). **Gerard Saucier:** Conceptualization (supporting), methodology (supporting), writing—review and editing (supporting). **Clemens M. Lechner:** Conceptualization (equal), data curation (supporting), formal analysis (equal), investigation (equal), methodology (lead), validation (supporting), visualization (supporting), writing—review and editing (lead).
Study 2 was published as Partsch, M. V., Bluemke, M., & Lechner, C. M. (2022). Revisiting the hierarchical structure of the 24 VIA character strengths: Three global dimensions may suffice to capture their essence. *European Journal of*	**Melanie V. Partsch:** Conceptualization (equal), data curation (lead), formal analysis (lead), investigation (lead), methodology (equal), validation (lead), visualization (lead), writing—original draft (lead), writing—review and editing (equal). **Matthias Bluemke:** Conceptualization (equal), data curation (supporting), formal

Personality, *36*(5), 825–845. https://doi.org/10.1177/08902070211017760 Note: The version included in this dissertation is the final author version before copy editing.	analysis (supporting), investigation (equal), methodology (equal), validation (supporting), writing—review and editing (supporting). **Clemens M. Lechner:** Conceptualization (equal), data curation (supporting), formal analysis (supporting), investigation (equal), methodology (equal), validation (supporting), writing—review and editing (equal).
Study 3 was informally published as Partsch, M. V., Olaru, G., & Lechner, C. M. (2022). *Measuring global character dimensions: An ant colony optimization approach towards three core strengths scales*. PsyArXiv. https://doi.org/10.31234/osf.io/tdjh4	**Melanie V. Partsch:** Conceptualization (equal), data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), software (equal), validation (lead), visualization (lead), writing—original draft (lead). **Gabriel Olaru:** Conceptualization (supporting), formal analysis (supporting), methodology (equal), software (equal), writing—review and editing (equal). **Clemens M. Lechner:** Conceptualization (equal), data curation (supporting), formal analysis (supporting), investigation (supporting), methodology (equal), validation (supporting), writing—review and editing (equal).

Mannheim, July 5, 2023 ______________________________

CURRICULUM VITAE

ACADEMIC EDUCATION

03/2017–12/2022	Doctoral Student, Psychology GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany RPTU Kaiserslautern-Landau, Germany
09/2013–10/2016	Master of Science in Psychology University of Mannheim, Germany
10/2010–09/2013	Bachelor of Science in Psychology University of Ulm, Germany

ACADEMIC POSITIONS

08/2023–present	Postdoctoral Researcher Utrecht University, Utrecht, the Netherlands Department of Methodology and Statistics Project “Detection and Classification of Misspecifications in Structural Equation Modeling Using Machine Learning”
03/2022–12/2022	Research Associate GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany Department Survey Design and Methodology Team Questionnaire Design and Evaluation
03/2022–09/2022	Research Associate University of Education, Weingarten, Germany Department of Educational Psychology Project “Individual and Contextual Determinants of Lifelong Learning”
03/2017–02/2022	Research Associate GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany Department Survey Design and Methodology Team Scale Development and Documentation

PEER-REVIEWED JOURNAL ARTICLES

Partsch, M. V., & Landberg, M. (2023). Modeling determinants of lifelong learning according to the Theory of Planned Behavior: A proxy-based approach using PIAAC data. *Adult Education Quarterly*. https://doi.org/10.1177/07417136231208951

Landberg, M., & Partsch, M. V. (2023). Perceptions on and attitudes towards lifelong learning in the educational system. *Social Sciences & Humanities Open*, *8*(1), Article 100534. https://doi.org/10.1016/j.ssaho.2023.100534

Partsch, M. V., Bluemke, M., & Lechner, C. M. (2022). Revisiting the hierarchical structure of the 24 VIA character strengths: Three global dimensions may suffice to capture their essence. *European Journal of Personality*, *36*(5), 825–845. https://doi.org/10.1177/08902070211017760

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

Lechner, C. M., Partsch, M. V., Danner, D., & Rammstedt, B. (2019). Individual, situational, and cultural correlates of acquiescent responding: Towards a unified conceptual framework. *British Journal of Mathematical and Statistical Psychology*, *72*(3), 426–446. https://doi.org/10.1111/bmsp.12164

Nießen, D., Partsch, M. V., Kemper, C. J., & Rammstedt, B. (2019). An English-language adaptation of the social desirability–gamma short scale (KSE-G). *Measurement Instruments for the Social Sciences*, *1*, Article 2. https://doi.org/10.1186/s42409-018-0005-1

ARTICLES UNDER REVIEW OR INFORMALLY PUBLISHED ON REPOSITORIES

Partsch, M. V., Schmidt, I., Haehner, P., & Lechner, C. M. (2023). *VaLiGo—Measuring basic human values and life goals with a nested 30–20–10-item inventory (German version)*. PsyArXiv. https://doi.org/10.31234/osf.io/evr7a

Stanciu, A., Bernardes, M., Partsch, M. V., & Lechner, C.M. (2023). *Basic human values and the adoption of cryptocurrency*. PsyArXiv. https://doi.org/10.31234/osf.io/hxutm

Partsch, M. V., Olaru, G., & Lechner, C. M. (2022). *Measuring global character dimensions: An ant colony optimization approach towards three core strengths scales*. PsyArXiv. https://doi.org/10.31234/osf.io/tdjh4

Bluemke, M., Partsch, M. V., Saucier, G., & Lechner, C. M. (2021). *Human character in the IPIP: Towards shorter, more content-valid, and cross-culturally comparable IPIP-VIA character strength scales*. PsyArXiv. https://doi.org/10.31234/osf.io/k79qf

Partsch, M. V., Danner, D., & Chernyshenko, O. (2018). *CCS-based 7-item traditionalism scale (CCS-T-7)*. Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/ZIS261

CONFERENCE PRESENTATIONS

Partsch, M. V., Schmidt, I., Haehner, P., Hanzlová, R., & Lechner, C. M. (2023, September 24–27). VaLiGo—Measuring basic human values and life goals with a nested 30–20–10-item inventory. In B. Rammstedt (Chair), *Assessment of personality, attitudes, and other constructs* [Symposium]. 17th Biennial Conference of the German Psychological Society - Personality Psychology and Psychological Diagnostics (DPPD) Section, Salzburg, Austria.

Partsch, M. V., Bluemke, M., Olaru, G., & Lechner, C. M. (2021, September 12–15). Identifying and measuring three global dimensions of the VIA trait space based on the IPIP-VIA-R. In M. Bluemke & L. Wagner (Chairs), *How can we best describe and measure character in relation to personality?* [Symposium]. 16th Biennial Conference of the German Psychological Society - Personality Psychology and Psychological Diagnostics (DPPD) Section, Ulm, Germany.

Partsch, M. V., Bluemke, M., & Lechner, C. M. (2019, September 16–18). Applying the Bass-ackwards approach to the IPIP-VIA-R inventory to establish the higher-order structure of Peterson and Seligman's 24 character strengths. In M. Bluemke (Chair), *Solutions to methodological challenges in the equivalent assessment of personality traits and values* [Symposium]. 15th Biennial Conference of the German Psychological Society - Personality Psychology and Psychological Diagnostics (DPPD) Section, Dresden, Germany.

Partsch, M. V., Danner, D., & Rammstedt, B. (2018, September 15–20). Der inkrementelle Nutzen einer 30 Items umfassenden RIASEC-Kurzskala in kulturübergreifenden Bevölkerungsumfragen [The incremental value of a 30-item RIASEC short scale in cross-cultural large-scale assessment]. In M. Bluemke (Chair), *Fortschritte der Skalen-entwicklung für internationale Vergleichsstudien (Large-Scale Assessment) [Progress in scale development for cross-national studies (large-scale assessment)]* [Symposium]. 51st congress of the German Psychological Society (DGPs), Frankfurt, Germany.

Partsch, M. V., Lechner, C. M., Bluemke, M., & Saucier, G. (2018, July 2–5). Character strengths: Towards a shorter (and more content-valid) measure of the IPIP-VIA scales. In D. Danner (Chair), *Advances in assessing non-cognitive skills in large-scale surveys* [Symposium]. 11th International Test Commission (ITC) Conference, Montréal, Canada.

Partsch, M. V., Danner, D., & Rammstedt, B. (2018, July 2–5). Vocational interests: The incremental value of a 30-item short scale suitable for large-scale assessment in six languages. In D. Danner (Chair), *Advances in assessing non-cognitive skills in large-scale surveys* [Symposium]. 11th International Test Commission (ITC) Conference, Montréal, Canada.

Partsch, M. V., Bluemke, M., Lechner, C. M., & Saucier, G. (2018, May 3–5). *Toward shorter (and more content-valid) measures of character strengths: Revising the IPIP-VIA scales* [Poster presentation]. 3rd conference of the German Society for Positive-Psychological Research (DGPPF), Bochum, Germany.

Partsch, M. V. (2017, September 4–6). *Facettentheoretische Entwicklung und Validierung einer Skala zu Personal Need for Structure [Applying facet theory to develop a Personal Need for Structure scale and its validation]* [Poster presentation]. 14th Biennial Conference of the German Psychological Society - Personality Psychology and Psychological Diagnostics (DPPD) Section, München, Germany.