Multitech Journal of Science and Technology (MJST)
Vol.3, No.8, 2026:1071-1098

Beyond the 16 Types: Hidden Subtypes, Personality Rarity, and the Psychometric Limits of MBTI-Based Classification in the Digital Age—A Critical Review

Loso Judijanto
IPOSS Jakarta
Corresponding Author: Loso Judijanto losojudijantobumn@gmail.com

ARTICLEINFO

Keywords: Myers–Briggs Type Indicator; MBTI; Personality Types; Personality Subtypes; Personality Rarity; Big Five; HEXACO; Psychometric Validity; Personality Profiles; Digital Identity

Received : 5 July
Revised : 20 August
Accepted: 20 September

©2026 Judijanto: This is an open-access article distributed under the termsof the Creative Commons Atribusi 4.0 Internasional.

ABSTRACT

Personality typologies have acquired renewed visibility through digital platforms, workplace applications, educational settings, and viral infographics that extend conventional personality classifications into increasingly granular "subtypes." One contemporary example ranks combinations derived from Myers–Briggs Type Indicator (MBTI) labels into categories ranging from "very common" to "ultra rare," including additional suffixes that appear to extend the conventional 16-type architecture. This qualitative literature review critically examines whether contemporary personality science supports such subtype and rarity claims. Literature published predominantly between 2020 and 2026 was synthesized across six domains: MBTI psychometrics, dimensional trait models, person-centered personality profiles, cross-cultural measurement, digital personality assessment, and applied personality classification. The synthesis shows that recent evidence provides a more nuanced picture of the MBTI than either uncritical endorsement or wholesale rejection. Contemporary studies report meaningful reliability and convergent-validity evidence for established MBTI dimensions, while important questions remain concerning categorical interpretation, structural validation, temporal stability, and predictive usefulness. More importantly, reliability of an established instrument does not establish the validity of derivative subtype systems or population-rarity rankings

INTRODUCTION

Personality classification occupies an unusual position at the intersection of scientific psychology, organizational practice, popular culture, and digital identity. Few psychological frameworks have achieved the public recognition enjoyed by the Myers–Briggs Type Indicator (MBTI). Its four preference dimensions—Extraversion–Introversion, Sensing–Intuition, Thinking–Feeling, and Judging–Perceiving—combine to produce 16 familiar four-letter personality designations. Although the MBTI originated in an earlier theoretical tradition than contemporary trait psychology, it continues to be used in counseling, education, professional development, healthcare education, leadership programs, and increasingly in digital social interaction (Ramachandran et al., 2020; Haidar et al., 2021; Li et al., 2024; Erford et al., 2025).

The contemporary digital environment has, however, transformed personality classification beyond conventional psychometric assessment. Four-letter designations increasingly function not merely as test results but as identity labels, conversational shorthand, dating preferences, memes, occupational stereotypes, social-media bios, and markers of perceived interpersonal compatibility. Wu et al. (2024), for example, documented how MBTI labels among Chinese young adults can become forms of social categorization linked with identity, belonging, impression management, and social interaction. Digital platforms have therefore created a context in which personality labels can acquire cultural meaning beyond the scientific claims of the instruments from which they originated.

The infographic that motivates the present review provides an especially useful illustration of this phenomenon. Instead of presenting only the conventional 16 MBTI labels, it combines them with suffixes such as AC, AH, OH, OC, and AN, creating labels such as INFJ-AC, INTJ-AC, ENFP-OH, and ESTP-AH. These combinations are then arranged in a pyramid allegedly representing rarity: approximately the top 60% are designated "very common," the top 40% "common," the top 25% "below average," the top 15% "somewhat rare," the top 5% "rare," and the top 1% "ultra rare." The highly visual presentation creates an impression of statistical precision, hierarchical structure, and population-based empirical knowledge.

Yet visual precision should not be mistaken for measurement precision. A statement that a psychological subtype belongs to the "top 1%" is considerably stronger than saying that a person resembles a descriptive personality category. A prevalence assertion requires an identifiable target population, representative sampling frame, validated measurement procedure, reproducible classification rule, denominator, sampling uncertainty, demographic adjustment, and preferably independent replication. The scientific question is consequently not merely whether personality types are useful. It is whether increasingly fine-grained derivative classifications have crossed the evidentiary distance between an intuitively appealing label and a validated population construct.

This distinction is especially important because recent evidence concerning the MBTI itself is more nuanced than polarized public debates imply. Erford et al. (2025), synthesizing 193 studies using MBTI Form M between 1999

and 2024, reported internal-consistency estimates ranging approximately from 0.845 to 0.921 and substantial convergent evidence across related personality instruments. At the same time, their synthesis identified an absence of structural-validity and test–retest studies in the reviewed Form M literature. Thus, contemporary evidence cannot appropriately be summarized either as "MBTI has no scientific evidence" or as "all MBTI-derived classifications are scientifically validated." The scientific validity of particular scores must be separated from the stronger ontology implied by discrete types and from still stronger claims concerning novel subtypes and population rarity.

Contemporary personality psychology also offers several alternatives to rigid typological reasoning. The Big Five and HEXACO traditions conceptualize personality primarily in dimensional terms, whereas emerging person-centered approaches use latent profile analysis and related methods to identify empirically recurring combinations of continuous traits. These approaches demonstrate that personality profiles can potentially be meaningful without assuming that every intuitively appealing classification corresponds to a naturally occurring psychological type. Research using very large HEXACO datasets, for instance, has identified reproducible profile structures under certain conditions, although reproducibility weakened considerably in smaller samples (Espinoza et al., 2020). Likewise, Kerber et al. (2021) derived personality prototypes using several clustering procedures and validated them against external outcomes.

The distinction has direct consequences for statements about rarity. If the underlying psychological characteristics are continuously distributed, a dichotomous label may transform a relatively small score difference near a threshold into a categorical difference. Adding further subtype boundaries multiplies this problem. A four-letter type can be subdivided once, twice, or many times, but mathematical subdivision alone does not demonstrate the existence of corresponding natural classes. Moreover, apparent rarity may result from classification rules, sample composition, demographic structure, language, culture, occupational selection, or participation bias rather than a universally rare psychological configuration.

Cross-cultural measurement adds a further challenge. Dong and Dumas's (2020) systematic review of personality-measure invariance found that full cross-cultural comparability was far from automatic. More recent studies of the Big Five and HEXACO similarly emphasize that some structural properties travel across populations better than others. García et al. (2022), for example, found configural and metric invariance for the HEXACO-60 across 18 countries but not scalar invariance. Thus, even for extensively researched trait models, directly comparing population means or prevalence across cultures requires methodological caution.

This issue is increasingly consequential because personality assessment has moved into high-stakes environments. Research now examines personality information in automated interviews, credit scoring, academic prediction, workplace performance, and algorithmic inference from digital behavior. Hickman et al. (2022) found that automated video-based personality assessments

varied considerably in reliability and validity depending on the criterion used for training. Woo and Sohn (2022) proposed an MBTI-based approach to alternative credit scoring, demonstrating how personality classifications can migrate into domains with direct material consequences. Such applications require a substantially stronger evidentiary threshold than recreational self-exploration.

Against this background, the present qualitative literature review addresses eight questions: What does contemporary evidence indicate about the reliability and validity of MBTI classification? How does categorical typing compare with dimensional personality science? Can derivative "hidden subtype" systems be scientifically justified? What is required before a subtype can legitimately be termed rare? How do empirically derived profiles differ from predetermined labels? How do personality labels interact with digital identity? What risks arise in high-stakes applications? And what evidence-based framework should guide future evaluation of personality-subtype claims?

The review argues that the correct scientific response is neither to dismiss all typological language nor to treat visually persuasive subtype systems as population facts. Instead, claims should be calibrated to evidence. An assessment can be useful for reflection without proving that it captures natural categories; a classification can show reliability without demonstrating predictive utility; and an established measure can possess evidence without automatically validating every derivative taxonomy built around its labels.

LITERATURE REVIEW

1. From Personality Description to Personality Classification

Personality psychology seeks to characterize relatively enduring patterns of thought, emotion, motivation, and behavior. Contemporary scholarship emphasizes that personality is broader than any single inventory, trait model, or self-report format. Roberts and Yoon (2022) describe personality psychology as a field concerned with meaningful individual differences that unfold across situations and the life course, while Rauthmann (2024) cautions against equating personality itself with self-reported Big Five scores. Personality therefore exists conceptually prior to any particular measurement system.

The dominant contemporary trait tradition generally represents personality dimensions as continuous. The Big Five organizes individual differences around Extraversion, Agreeableness, Conscientiousness, Negative Emotionality/Neuroticism, and Openness. HEXACO adds Honesty–Humility while reorganizing portions of the remaining structure. Research continues to refine both broad domains and narrower facets, demonstrating that substantial predictive information may reside below the domain level (Strus and Cieciuch, 2021; Stewart et al., 2022; Irwing et al., 2024).

Bainbridge et al. (2022) examined whether commonly used psychological trait scales can be meaningfully organized within the Big Five and found substantial but not complete coverage. Feher and Vernon (2021) likewise reviewed alternative personality models and constructs extending beyond the Big Five. These studies matter for evaluating MBTI extensions because they demonstrate that personality complexity cannot be solved simply by multiplying

categorical labels. Additional psychological nuance requires theoretical definition and psychometric demonstration, not merely nomenclature.

Mõttus et al. (2020) offer an especially useful distinction between description, prediction, and explanation. A personality construct may describe individual differences effectively without strongly predicting a particular outcome, and a predictor may work without explaining underlying causal mechanisms. This framework is directly relevant to MBTI debates. A four-letter label may provide users with an understandable descriptive narrative, but descriptive usefulness does not imply that the same label should predict leadership, academic success, occupational fit, credit default, or relationship compatibility.

2. Contemporary Evidence Concerning MBTI

The MBTI remains widely used, and recent research does not support caricaturing it as either completely evidence-free or conclusively established. Erford et al. (2025) reported relatively strong internal consistency and convergent evidence for Form M while identifying important gaps concerning structural validity and temporal stability within the studies available to their synthesis. This is a significant recent contribution because it moves discussion from generalized criticism toward specific evidentiary domains.

Nugraha et al. (2025), studying more than 7,500 Indonesian students, reported evidence supporting the measurement structure of MBTI dimensions in their sample. Such results demonstrate why broad declarations that MBTI dimensions contain no measurable information are difficult to sustain. At the same time, demonstrating adequate measurement of a dimension is conceptually distinct from demonstrating that a median split or preference boundary reveals two naturally separate kinds of people.

Predictive evidence is similarly context-dependent. Zárate-Torres and Correa (2023) found satisfactory measurement characteristics for the personality and leadership constructs they studied but weak relationships between MBTI variables and leadership practices. The result illustrates the critical distinction between measuring something consistently and predicting an outcome strongly. Healthcare education provides further examples. Ramachandran et al. (2020) reviewed MBTI use in medical education, while Haidar et al. (2021) reported distributions of MBTI types among ophthalmology residents. Li et al. (2024) subsequently synthesized decades of MBTI-based research among dental students. These studies can illuminate patterns within specific professional samples, but occupational prevalence cannot automatically be generalized to the population because selection into medicine, dentistry, or a medical specialty may itself be associated with personality and educational background.

Choong and Varathan (2021) further illustrate the migration of MBTI labels into computational contexts by predicting the Judging–Perceiving dimension from online forum data. Such research demonstrates that linguistic or behavioral traces may contain information associated with self-reported personality labels, but it does not independently establish the ontological reality of all four-letter categories, still less of additional subtype suffixes.

3. Dimensional Traits, Facets, and Nuances

Trait models offer an important scientific comparator because their psychometric development increasingly operates at multiple hierarchical levels. Stewart et al. (2022), analyzing prediction across Big Five domains, facets, and individual items, found that narrower information could improve prediction of life outcomes. Irwing et al. (2024) similarly developed a taxonomy of personality facets based on broad item pools. These findings support the idea that personality contains meaningful fine-grained variation, but they simultaneously show what rigorous granularity looks like: systematic construct mapping, item development, replication, and criterion validation.

The implication is important for so-called hidden subtypes. The scientific objection is not that personality must remain broad. Contemporary research explicitly investigates facets and even item-level nuances. The objection is that finer categories require more evidence, not less. When a popular system adds suffixes to an existing four-letter label, researchers must determine whether those suffixes correspond to stable latent constructs, distinct locations in multidimensional trait space, or merely an arbitrary repackaging of already continuous characteristics. A large taxometric meta-analysis further supports this caution: dimensional solutions substantially outnumbered taxonic solutions across psychological constructs, indicating that categorical discontinuities should be demonstrated rather than presumed (Haslam et al., 2020).

Zell and Lesick's (2022) synthesis of more than 50 meta-analyses, encompassing more than half a million participants, confirms that broad personality traits exhibit reproducible but generally modest associations with performance. Conscientiousness produced the strongest overall relationship, but effect sizes varied substantially by criterion. Mammadov's (2022) meta-analysis similarly showed meaningful personality associations with academic performance. These literatures demonstrate how credible predictive claims emerge: through large cumulative evidence bases rather than isolated labels.

Personality is also not perfectly immutable. Graham et al. (2020), coordinating evidence across 16 longitudinal samples involving more than 60,000 participants, identified both stability and systematic personality change across the life span. Genetic research further indicates substantial biological contributions while preserving an important role for environment and development (Gupta et al., 2024). Consequently, a subtype label should not automatically be treated as a timeless essence.

4. Person-Centered Profiles as a Scientifically Relevant Middle Ground

Typological thinking persists partly because people naturally understand configurations more easily than vectors of continuous scores. Person-centered methods therefore offer an instructive middle ground between purely dimensional models and predetermined folk typologies.

Espinoza et al. (2020) analyzed four extremely large international HEXACO samples and identified a reproducible five-profile structure. Crucially, the authors found that profile replication became unreliable when samples fell below approximately 500 respondents. This finding is highly relevant to rarity claims: even empirically derived clusters can be sensitive to sample size, and stable profile identification requires far more than intuitive visual categorization.

Kerber et al. (2021) used latent profile analysis, k-means clustering, and spectral clustering on more than 22,000 participants and identified five prototypes that could be evaluated against behavioral and well-being criteria. Again, the critical scientific feature is not the existence of labels such as "resilient" or "overcontroller," but the inferential chain by which profiles were extracted from data and externally validated.

Smaller person-centered studies reinforce the usefulness—but also context-dependence—of such approaches. Ratchford et al. (2022) identified personality profiles differentially associated with well-being. Exley et al. (2022) linked Big Five profiles with income, risk tolerance, and life satisfaction. Preston et al. (2022) analyzed family personality configurations and later health and happiness. Kokkinos et al. (2023) connected student personality profiles with procrastination and academic dishonesty. Lan et al. (2021) identified employee profiles related to job satisfaction and performance in hospitality settings.

Collectively, this literature demonstrates that "personality profiles" can be scientifically meaningful. It does not, however, imply that any predetermined subtype taxonomy is valid. Empirically generated profiles are estimated from patterns in continuous data, accompanied by fit statistics, probability-based assignment, sensitivity analyses, external validation, and ideally replication. In contrast, many popular subtype systems begin with predefined labels and subsequently treat their existence as established.

METHODOLOGY

1. Review Design

This study employs a critical qualitative literature review, not a systematic literature review or meta-analysis. The objective is theoretical integration rather than exhaustive enumerative retrieval. The review synthesizes contemporary evidence addressing MBTI psychometrics, dimensional personality models, person-centered personality profiles, measurement invariance, personality assessment methodology, digital personality prediction, and applied classification.

The distinction from a systematic literature review is deliberate. A systematic review typically requires a predetermined protocol, comprehensive database search, reproducible inclusion and exclusion procedures, formal study-screening counts, and often structured risk-of-bias assessment. The present review instead uses iterative purposive searching to identify theoretically informative evidence necessary to evaluate a central conceptual problem: what must be demonstrated before a derivative personality subtype can legitimately be described as rare?

2. Literature Identification

Priority was given to peer-reviewed publications between 2020 and 2026, with emphasis on journals in personality psychology, psychological assessment, organizational psychology, educational psychology, behavioral science, and computational personality research. Search concepts included combinations of *MBTI*, *Myers-Briggs*, *personality type*, *personality profile*, *personality subtype*, *personality rarity*, *Big Five*, *HEXACO*, *latent profile analysis*, *measurement invariance*,

psychometric validity, *digital personality*, *personality prediction*, and *automated personality assessment*.

The evidence base was intentionally broadened beyond studies directly naming the MBTI because evaluation of hidden subtypes requires standards derived from contemporary psychometrics and personality science. For example, studies validating the BFI-2 across Japan, Turkey, Brazil, Bangladesh, and Korea provide an important benchmark for the work required before asserting cross-cultural comparability of a personality construct (Cemalcilar et al., 2021; Yoshino et al., 2022; Pires et al., 2023; Siraji et al., 2025; Choi et al., 2026).

3. Analytical Strategy

Six thematic lenses guided the analysis. These were psychometric evidence (reliability, structural validity, convergent and discriminant validity, temporal stability, and predictive validity); personality ontology (dimensional versus categorical representation); profile science (person-centered clustering and latent profile analysis); population inference (requirements for estimating prevalence and rarity); cross-cultural transportability (invariance across language, culture, sex, age, and population); and applied consequences (digital interaction, education, employment, leadership, finance, and algorithmic assessment).

Contradictory evidence was treated as substantively informative rather than as noise. This is especially appropriate for the MBTI, where different studies answer different questions. Reliability studies, factor-analytic studies, predictive studies, prevalence studies, and sociocultural studies should not be treated as interchangeable forms of evidence.

RESULTS

1. Theme 1: The Scientific Status of Standard MBTI is Different From the Status of Derivative Subtype Systems

The most important finding is conceptual. Evidence concerning the MBTI cannot simply be transferred to every system that incorporates MBTI letters. A derivative label may contain "INTJ" or "INFJ," but this does not mean that an appended suffix has inherited the reliability, validity, normative database, or interpretive status of the underlying instrument.

Recent MBTI research supports this distinction. Erford et al. (2025) provide evidence concerning MBTI Form M. Nugraha et al. (2025) evaluate an MBTI factor structure in Indonesian students. Zárate-Torres and Correa (2023) analyze relationships with leadership behavior. None of these studies, however, automatically validates arbitrary additional codes attached to four-letter types. A new suffix constitutes an additional measurement claim requiring separate construct definition and validation.

This matters for the motivating infographic because labels including AC, AH, OH, OC, and AN are displayed as though they belong to a unified population taxonomy. Within the recent peer-reviewed literature examined for this review, no validation evidence was identified establishing these particular suffixes as standardized extensions of MBTI Form M or as population categories whose prevalence corresponds to the percentages shown in the infographic. The appropriate academic treatment is therefore to regard the infographic as a case illustration of digital subtype proliferation, not as a population dataset.

A useful general principle follows from the distinction between validated parent measures and unvalidated derivative taxonomies. It can be stated compactly as the following evidentiary inequality. The compact expression below summarizes this non-transferability of evidence.

Evidence(parent construct) ≠ Evidence(derivative subtype)

The principle applies beyond MBTI. A validated depression inventory does not validate every internet quiz incorporating the word "depression"; a validated Big Five measure does not validate every social-media category named after one of its traits. Scientific inheritance must be demonstrated, not assumed.

2. Theme 2: Reliability is Necessary But Insufficient

Psychological classification requires reliable measurement, but reliability is only the beginning. Erford et al. (2025) reported substantial internal-consistency evidence for MBTI Form M, while also identifying important gaps in the contemporary literature they synthesized. This juxtaposition shows why the frequent question "Is MBTI scientifically valid?" is too broad. Different forms of validity answer different questions.

Reliability indicates whether a measurement procedure produces sufficiently coherent scores. Structural validity asks whether the expected latent organization is supported. Convergent validity asks whether the measure corresponds appropriately with related constructs. Discriminant validity asks whether it remains distinguishable from different constructs. Test–retest reliability concerns temporal stability. Criterion validity asks whether scores relate meaningfully to relevant outcomes.

Modern Big Five research illustrates this multidimensional approach. Ober et al. (2021) examined within-domain factor structure in adolescents; Gallagher et al. (2023) developed and validated a shortened Big Five Aspects Scale across several studies; Ruchensky et al. (2025) developed a response-consistency indicator for the BFI-2; Kang et al. (2024) validated an abbreviated Big Five measure in very large samples while examining retest reliability and external associations.

Response format also matters. Wetzel and Frick (2020) demonstrated that forced-choice versus rating-scale formats can alter aspects of reliability and construct validity. Pavlov et al. (2021) showed that even apparently technical details such as desirability matching influence forced-choice assessment design. Otero et al. (2020) found substantial convergent and predictive evidence across single-stimulus and quasi-ipsative measures, again emphasizing that measurement architecture is part of the construct-validation problem.

Thus, even if a subtype questionnaire produced internally consistent responses, that fact alone would not demonstrate that the subtype represents a distinct psychological category. Nor would it establish population prevalence.

As illustrated in Figure 1, a rarity statement sits at the top rather than the bottom of the evidentiary hierarchy. A subtype cannot legitimately become a population-prevalence statistic merely because its label has been defined or because a questionnaire can assign respondents to it. Each inferential level introduces additional assumptions that must be tested. Figure 1 therefore

operationalizes the distinction between defining a construct and making a defensible population-prevalence claim.

Figure 1. The Personality-Subtype Evidence Ladder

Figure 1 illustrates that the evidence ladder also resolves an important misunderstanding in popular discussions. Scientific criticism of a subtype does not necessarily mean that the parent personality assessment is entirely useless. A system may satisfy Levels 1–2 while lacking Level 5 or Level 7 evidence. Claims should therefore be restricted to the highest level actually supported.

3. Theme 3: Continuous Personality Variation Complicates Categorical Rarity

A central theoretical difficulty concerns the relationship between continuous variation and discrete categories. Most contemporary trait models assume that individuals vary by degree rather than belonging exclusively to naturally separate kinds. Big Five and HEXACO research therefore usually retains continuous scores.

This does not prohibit categories, but it changes their interpretation. Suppose two respondents score 49 and 51 around an arbitrary threshold of 50. A dichotomous classification may place them into opposite categories even though they are psychologically similar. Conversely, two people scoring 51 and 95 may receive the same categorical label despite substantial differences. Every additional binary division creates another threshold at which continuous variation becomes categorical difference.

Sleep et al. (2021), comparing very brief Big Five measures, underscore the loss of psychometric information that can accompany aggressive measurement compression. Stewart et al. (2022) show the opposite side of the problem: narrower facets and items may contain useful predictive information, but preserving nuance is not equivalent to multiplying coarse binary labels.

This produces an apparent paradox. Popular personality culture often becomes "more precise" by adding categorical subdivisions, whereas contemporary personality science often becomes more precise by measuring continuous narrower characteristics. The former multiplies labels; the latter improves measurement resolution.

Marengo et al. (2021), for example, demonstrated systematic relationships between primary emotional systems and Big Five dimensions without needing to convert each participant into one immutable personality species. Similarly, Strus and Cieciuch (2021) investigated higher-order relations between HEXACO and personality metatraits using dimensional structural analysis.

Consequently, a 64-type or 80-type scheme may appear more sophisticated than a 16-type scheme while actually containing less information than well-measured multidimensional trait scores. Greater label granularity should therefore not be equated with greater psychometric information.

4. Theme 4: Empirically Derived Profiles Show That Types Can Be Useful—But Must Emerge from Data

The person-centered literature prevents the argument from becoming an overly simplistic "dimensions good, types bad" dichotomy. Empirical clustering demonstrates that recurring trait configurations can be identified under suitable conditions.

Espinoza et al. (2020) identified a replicable five-profile HEXACO solution across four large international samples of approximately 90,000 respondents each. Importantly, the authors reported that the structure could not be consistently rediscovered using samples below about 500 respondents. The implication for rarity estimates is profound. If even data-driven personality profiles require large samples to stabilize, an ultra-rare subtype representing around 1% of a population would require particularly strong sample sizes to estimate with acceptable precision.

Kerber et al. (2021) reached a related conclusion from a large representative German dataset, identifying five broad prototypes through multiple clustering techniques. Ratchford et al. (2022) found that trait-level and facet-level latent profile analyses could produce somewhat different solutions, with additional nuance appearing at the facet level. Thus, the number of profiles depends partly on the measurement resolution and statistical model.

This finding undermines the idea that there must be one universally correct number of personality categories. A profile system is partly an empirical model—a simplified representation chosen because it captures meaningful patterns. It should therefore be evaluated according to fit, replication, interpretability, predictive utility, and robustness rather than treated as a metaphysical inventory of human kinds. More recent person-centered studies also show that profile solutions can vary across social and educational contexts,

reinforcing the need to test profile stability and external validity rather than assume universal types (Yang and Xu, 2024; Kokkinos et al., 2024).

Figure 2 distinguishes psychological granularity from categorical proliferation. Narrower personality information can become increasingly detailed while remaining continuous, whereas subtype systems transform combinations of characteristics into named categories. The figure emphasizes that higher informational resolution can be achieved through validated continuous facets and nuances without multiplying unvalidated categorical labels.

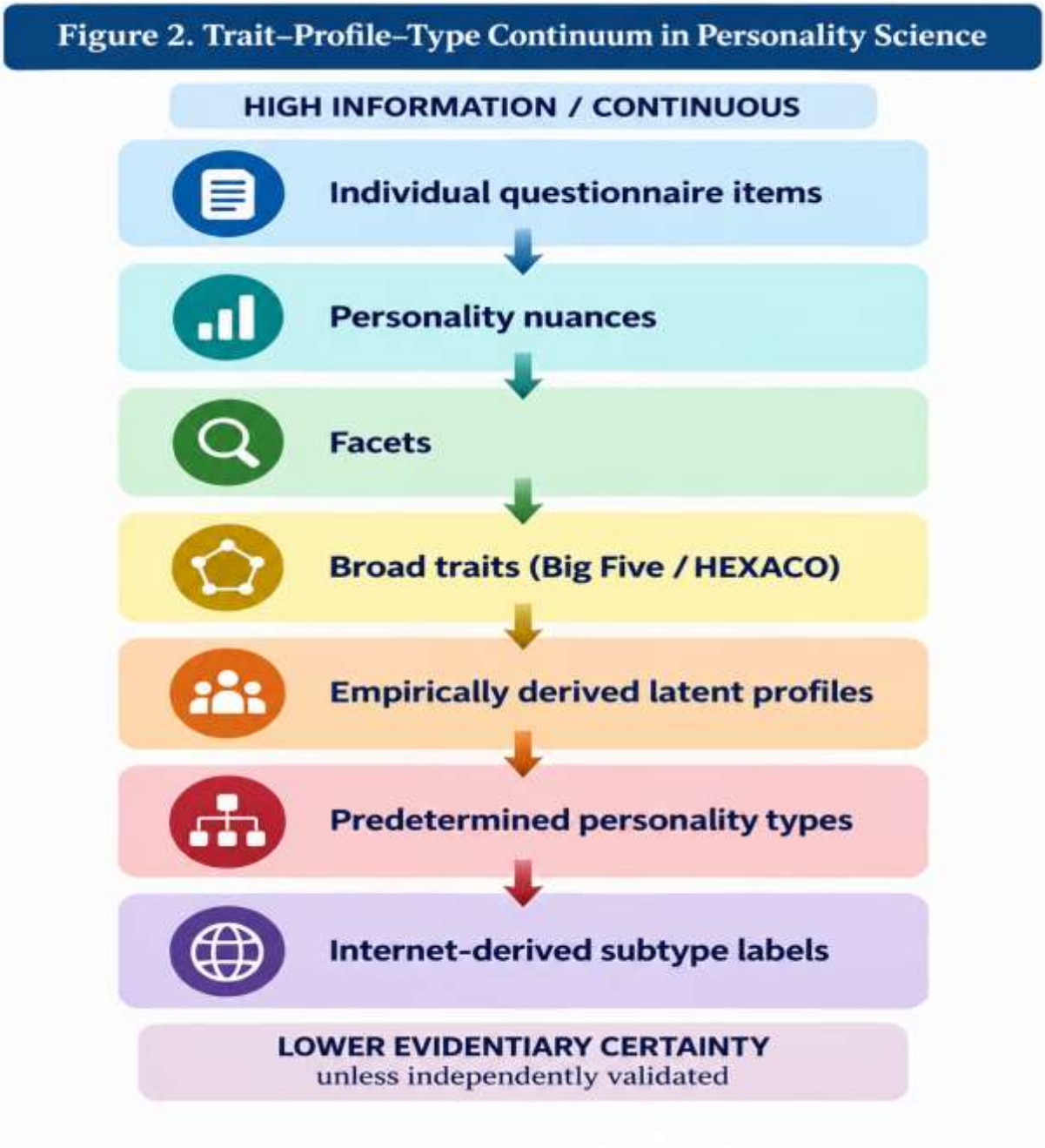

Figure 2. Trait–Profile–Type Continuum in Personality Science

Figure 2 illustrates that the continuum explains why an Internet subtype containing additional letters should not automatically be considered more precise than a standard personality inventory. Precision depends on measurement quality and information retained, not on the number of letters in the label.

5. Theme 5: Rarity is Fundamentally a Population-Estimation Problem

The strongest weakness in many online personality rarity charts is not necessarily the personality description itself but the absence of a transparent denominator. To say that a subtype represents 1% of people, researchers must specify the target population, explain how the sample was selected, disclose whether participation was self-selected, and describe the distributions of sex, age, education, occupation, language, and country. They must also establish whether the measure is invariant across those groups, state the rule used to generate the subtype, and quantify uncertainty around the prevalence estimate.

Consider a simple illustrative case. If an online personality website estimates a subtype from voluntary users, its prevalence describes—at best—the site's participating users under that site's classification method. Users attracted to personality testing are unlikely to constitute a random population sample. If different personality characteristics influence willingness to take such tests or share results, selection bias can directly alter the apparent distribution.

Occupational samples create another problem. Haidar et al. (2021) found particular MBTI patterns among ophthalmology residents, and Li et al. (2024) examined dental-student type distributions. Such findings do not establish the prevalence of those types among all adults. Medical training populations differ from general populations in age, education, career selection, and possibly personality-related characteristics.

Prevalence estimates are also sensitive to subgroup composition. Graham et al. (2020) demonstrated age-related personality changes, while García et al. (2022) identified gender, age, social-position, and national differences within HEXACO research. A population-rarity ranking that does not describe its demographic weighting therefore lacks crucial interpretive information.

A "top 1%" statement should additionally provide statistical uncertainty. An observed prevalence of 1% in a sample of 500 corresponds to only approximately five people. Small changes in classification can dramatically change the estimated prevalence. Rare-category research consequently needs very large samples, especially when estimating differences across countries, age groups, or sexes. For a sample of size n with r classified cases, the sample prevalence is $\hat{p} = r/n$ and its binomial standard error is $SE(\hat{p}) = \sqrt{[\hat{p}(1 - \hat{p})/n]}$. An approximate 95% interval is $\hat{p} \pm 1.96 \times SE(\hat{p})$, although exact or Wilson-type intervals are preferable when prevalence is very low or samples are modest.

6. Theme 6: Cross-Cultural Measurement Challenges Universal Rarity Rankings

Cross-cultural personality research provides perhaps the strongest reason to avoid universal rarity hierarchies without evidence. Dong and Dumas (2020) reviewed measurement-invariance studies across cultural, gender, and age groups and found substantial limitations in the extent to which personality scales

achieved strong forms of cross-cultural invariance. Scalar invariance, necessary for straightforward comparison of means, was especially difficult across cultural groups.

Thalmayer et al. (2022) similarly argue that personality research must move beyond simply imposing Western structural models in every cultural context and should integrate etic and emic perspectives. Personality language may reflect universally important human differences while still being shaped by cultural environments.

Recent validation studies demonstrate the work required to establish transportability. Cemalcilar et al. (2021) tested the BFI-2 in a Turkish non-WEIRD community sample and reproduced many—but not all—psychometric and trait–outcome relationships. Yoshino et al. (2022) validated a Japanese adaptation while explicitly examining measurement invariance. Pires et al. (2023) investigated the Brazilian BFI-2 and reported adequate results for many facets alongside limitations requiring further validation.

Siraji et al. (2025) found substantial support for a Bangla BFI-2 but only weaker invariance across languages, illustrating that linguistic differences remain relevant even when a broad model performs well overall. Choi et al. (2026) validated the Korean BFI-2 and explicitly compared its properties with English and Chinese versions. Cupani et al.'s (2025) Latin American research similarly investigated factor structure and behavior across regional populations rather than assuming equivalence a priori.

If such cross-cultural testing is necessary for the Big Five—one of the world's most extensively researched personality frameworks—it is difficult to justify universal rarity statements for a newer subtype scheme without comparable validation. This concern is reinforced by large-scale evidence showing that measurement invariance can differ substantially across personality facets and demographic groups, as well as by a recent cross-cultural review that found full scalar invariance to be uncommon across personality instruments (Temizyürek et al., 2024; Sheppard et al., 2026).

Figure 3 illustrates why the statement “this subtype is 1% of the population” contains far more inferential content than a questionnaire result. Every stage creates a potential source of bias or uncertainty. It maps the sequence from raw responses to validated measurement, explicit classification, replication, invariance, representative sampling, weighting, and uncertainty estimation.

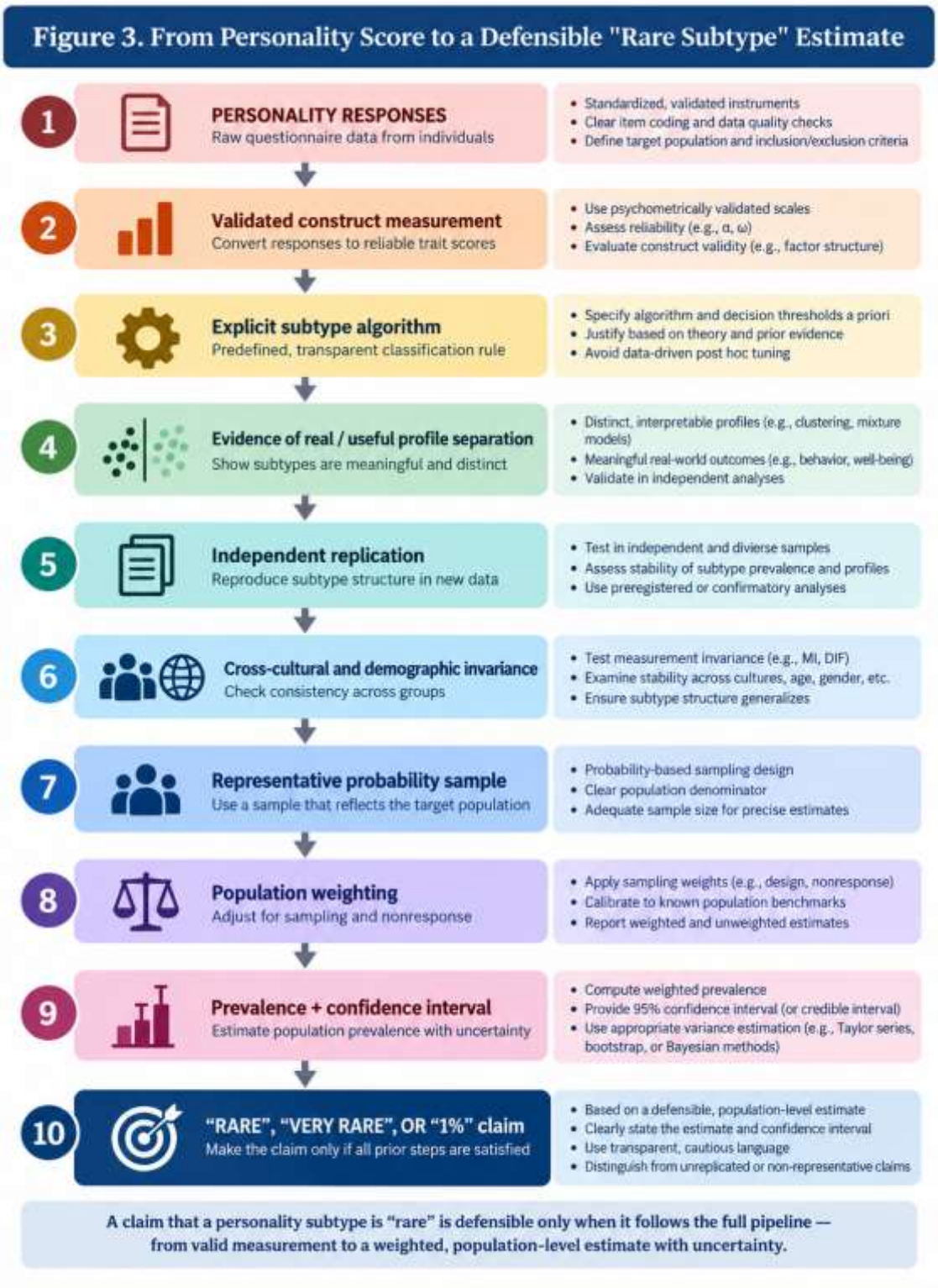

Figure 3. From Personality Score to a Defensible "Rare Subtype" Estimate

Figure 3 illustrates that most viral rarity graphics display only the final claim while concealing the intermediate analytical steps. Scientific reporting should reverse this pattern: methodology should be more visible than the label.

7. Theme 7: Personality Labels Can Become Social Identities

The popularity of MBTI cannot be understood solely through psychometrics. Labels provide a vocabulary through which people narrate themselves. A four-letter type is cognitively efficient: it converts complex, probabilistic information into an easily remembered social identity.

Wu et al. (2024) examined precisely this transition, describing MBTI use among Chinese youth as increasingly functioning as a social label rather than merely an assessment outcome. Their study found a more complex relationship with social anxiety than a simple direct effect, with identity and impression-management processes playing meaningful roles.

Personality categorization can also affect intergroup perceptions. Prinyapol et al. (2023) studied multiple personality categorization in relation to attitudes between Buddhist and Muslim students, showing how personality-based categorization can intersect with broader social-identity processes.

These findings suggest both benefits and risks. Labels may facilitate self-reflection, communication, normalization of interpersonal differences, and social belonging. Someone who learns that others share similar preferences may experience relief or greater self-understanding.

However, the same labels can produce essentialism: the belief that category membership defines what a person fundamentally is. Such essentialism can encourage people to interpret ordinary behavior through the label, selectively notice confirming evidence, overlook within-type variation, assume compatibility or incompatibility, constrain career or relationship choices, treat category boundaries as biologically fixed, and assign status based on perceived rarity.

The final risk is especially relevant to rarity infographics. A ranking from "very common" to "ultra rare" does not merely describe prevalence; it can acquire implicit status meaning. Scarcity is often socially associated with uniqueness or value. An "ultra rare" label may therefore create psychological incentives to identify with or preserve the classification even when measurement uncertainty is considerable.

8. Theme 8: Digital Personality Inference Makes Validation More—Not Less—Important

Contemporary personality assessment increasingly occurs without conventional questionnaires. Digital footprints, language, transactions, and video behavior can be analyzed computationally to estimate personality characteristics.

Hall and Matz (2020) showed that item-level personality nuances can modestly improve prediction from digital footprints. Alexander et al. (2020) reviewed both opportunities and problems created by big-data personality measurement, emphasizing psychometric reliability, validity, fairness, and privacy.

Hickman et al. (2022) provide an especially relevant cautionary example. Their automated video-interview models showed substantially different psychometric quality depending on how personality targets were operationalized. Models trained against interviewer ratings performed better on several validity criteria than those trained against self-reports. The finding

demonstrates that algorithmic sophistication cannot compensate for an unclear criterion.

Stevenor et al. (2024) extended this problem into high-stakes interviewing and found that automated personality estimates could behave differently depending on training context and criterion. Machine-learning systems therefore do not eliminate the classical psychometric questions; they amplify them. Independent auditing of commercial AI personality-prediction systems has likewise found instability across input formats and sources, illustrating why algorithmic outputs require direct reliability and validity evidence before high-stakes use (Rhea et al., 2022).

Choong and Varathan (2021) similarly demonstrate that MBTI-related labels can be computationally inferred from online language. But algorithmic predictability should not be confused with construct truth. A model may learn stylistic signals associated with how users self-identify without proving that the underlying categorical taxonomy represents natural psychological divisions.
Figure 4 presents personality labeling as a potential feedback process rather than a neutral one-time measurement event. The figure highlights how repeated self-identification, social comparison, stereotype expectations, and selective interpretation may reinforce an initially descriptive label.

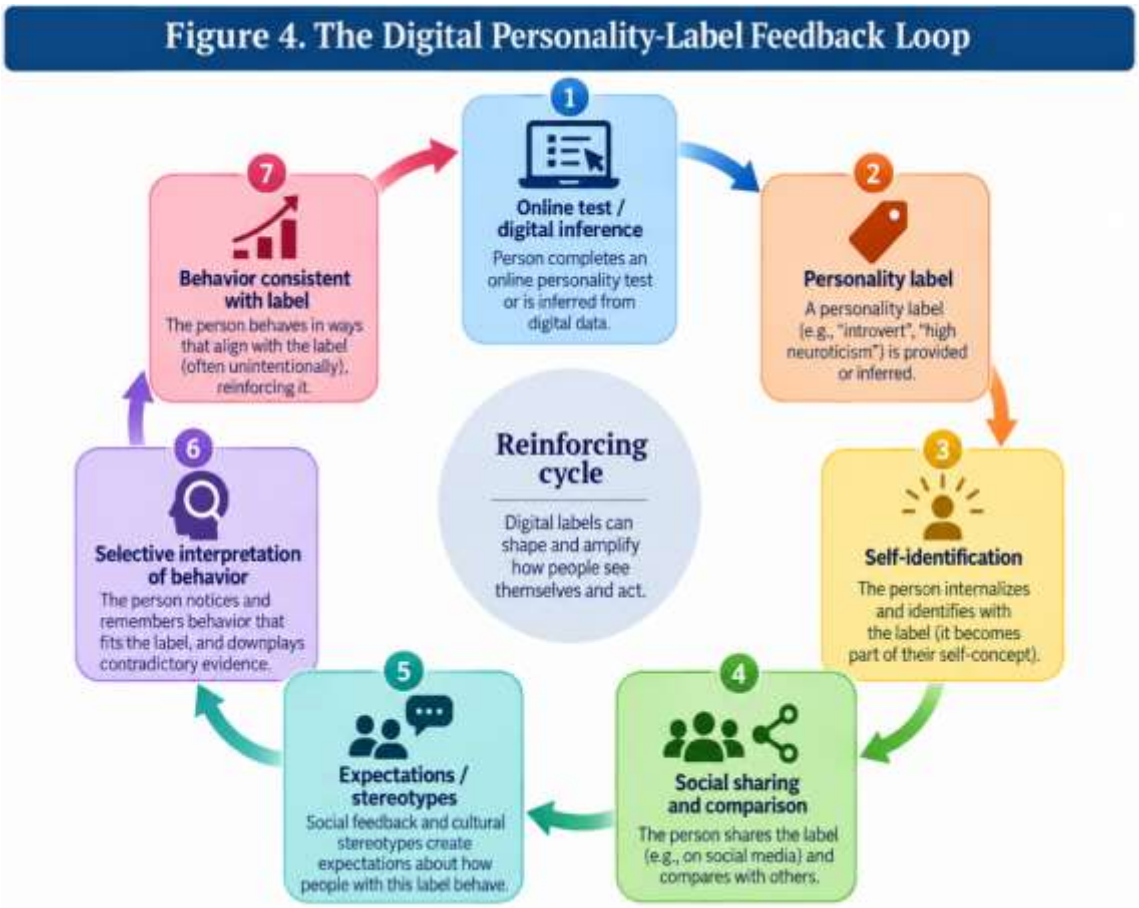

Figure 4. The Digital Personality-Label Feedback Loop

Figure 4 illustrates that the framework does not imply that every user becomes trapped by a personality label. Rather, it identifies a pathway through which descriptive categories can become increasingly self-confirming. This possibility is particularly important when labels are framed as rare, elite, or highly distinctive.

9. Theme 9: The Evidence Threshold Should Rise With Decision Stakes

Personality testing can serve different purposes, and the required scientific threshold should vary accordingly. Using a personality questionnaire as a conversation starter in a voluntary workshop is fundamentally different from rejecting an employment applicant because of a type. Similarly, discussing a personality profile in counseling differs from altering a person's credit risk based on inferred personality.

Woo and Sohn (2022) demonstrate that MBTI-derived information has been investigated in alternative credit-scoring models. Zárate-Torres and Correa (2023) studied leadership prediction and found comparatively weak personality–leadership relationships despite adequate measurement properties. These examples show why high-stakes applications must focus on incremental predictive validity rather than intuitive face validity.

The broader trait literature offers a useful benchmark. Zell and Lesick (2022) synthesized 54 meta-analyses and found modest but reproducible performance associations. Mammadov (2022) synthesized 267 samples concerning academic performance. Such evidence bases permit calibrated statements regarding average effects. In contrast, assigning a managerial, educational, or financial consequence to a narrowly defined hidden subtype without equivalent validation would represent a much larger evidentiary leap.
The same principle applies to very short assessments. Sleep et al. (2021) found that brevity can reduce measurement quality. Kang et al. (2024) therefore developed a new abbreviated inventory using extremely large samples rather than merely deleting items for convenience. Responsible classification requires demonstrating that convenience has not overwhelmed measurement.

DISCUSSION

1. How Rare is a "Hidden Subtype"? The Scientifically Appropriate Answer

The question posed by the infographic—"How rare is your hidden subtype?"—cannot be answered scientifically from the infographic alone. The scientifically appropriate answer is that rarity remains unknown unless the subtype definition, classification mechanism, sampling population, and prevalence methodology are independently documented and validated.

This is not semantic caution. It follows directly from population inference. A label such as *INFJ-AC* contains at least two levels of classification: assignment to INFJ and assignment to AC. Each may contain measurement error. The joint prevalence of the combined category therefore depends on the reliability and dependence structure of both classifications.

If AC is not an independently validated construct, estimating *INFJ-AC prevalence* produces numerical precision without demonstrated construct validity. A database containing a million responses cannot solve this problem. Large sample size reduces sampling error but does not repair construct

invalidity, self-selection, nonrepresentative recruitment, or an arbitrary classification algorithm.

This distinction between big data and good measurement is central to contemporary psychometrics. Alexander et al. (2020) emphasize that machine-learning and large-scale data create powerful opportunities but do not remove traditional requirements of reliability and validity.

2. Why "Rare" Does Not Mean "Better"

Rarity is descriptive, not evaluative. Even when prevalence is established accurately, low frequency does not imply greater intelligence, creativity, moral value, leadership potential, originality, or psychological depth. Human characteristics vary in adaptive value across environments. Some relatively common characteristics are highly functional; some uncommon configurations may be neutral, beneficial, disadvantageous, or context-dependent.

This becomes important when personality graphics use pyramid layouts. Visual hierarchies often carry unintended evaluative meaning: lower or narrower tiers can look like higher achievement levels, elite rankings, or scarce resources. Personality science provides no general principle according to which a 1% personality configuration is more desirable than a 20% configuration.

Indeed, broad performance evidence favors specific trait–criterion relationships rather than rarity itself. Zell and Lesick (2022) found that conscientiousness was consistently related to performance, whereas other Big Five relationships varied by outcome. The scientifically meaningful question is therefore not "How rare is the trait?" but "How is this trait related to the particular behavior or outcome under consideration?"

3. A Seven-Level Evidentiary Standard for Future Subtype Research

`The Personality Subtype Evidence Ladder proposed in this review provides a practical framework. Its levels are cumulative, so higher-order claims should be made only after the lower evidentiary requirements have been met.

Level 1: Construct Definition

Researchers must explain exactly what the subtype represents. If "AC" is proposed as a psychological dimension or category, its theoretical content must be stated without circularity. A category cannot be defined merely as "people whose test says AC."

Level 2: Reliable Measurement

Items or behavioral indicators must measure the construct with acceptable precision. Internal consistency, response consistency, measurement error, and score reliability should be assessed. Studies such as Gallagher et al. (2023) and Ruchensky et al. (2025) illustrate contemporary expectations.

Level 3: Structural Validity

Researchers must establish whether the proposed structure fits the data. Does AC represent a separate latent dimension? Is it reducible to existing traits? Are categories genuinely discrete? Does factor analysis, mixture modeling, or another appropriate method support the structure?

Level 4: Temporal and Sample Replication

The same individuals should be assessed over time, and the structure should replicate in independent samples. Espinoza et al. (2020) demonstrate that apparent profile solutions can depend on sample size.

Level 5: Cross-Cultural and Demographic Invariance

The system must be tested across language, country, age, sex, education, and relevant demographic contexts. Dong and Dumas (2020), Cemalcilar et al. (2021), Yoshino et al. (2022), Pires et al. (2023), Siraji et al. (2025), and Choi et al. (2026) collectively demonstrate the level of work required.

Level 6: External and Predictive Validity

Subtype membership should explain meaningful outcomes beyond existing measures. If the subtype does no better than a person's Big Five scores, established MBTI preferences, or demographic characteristics, its incremental utility may be limited.

Level 7: Representative Prevalence

Only after the preceding steps should claims such as "top 1%" be entertained. Prevalence studies require representative or appropriately weighted population samples, transparent denominators, uncertainty intervals, and replication.

4. Typologies as compression devices

Why, despite these limitations, do typologies remain attractive? A major answer is cognitive compression. Human personality is multidimensional. A complete personality profile containing dozens of continuous scores is difficult to remember and discuss. Labels reduce this informational burden.

A four-letter MBTI code is therefore analogous to a compressed file: it is easy to transmit but necessarily omits detail. Compression is not inherently problematic. Scientific models always simplify reality. The critical question is whether users remember that information has been compressed.

Problems arise when cognitive compression is progressively reified and then essentialized: compression → reification → essentialism. Compression means that a label summarizes tendencies; reification treats the label as a distinct thing; and essentialism treats the label as what the person fundamentally is. Digital subtype culture can accelerate this progression because labels often circulate without the technical documentation needed to preserve uncertainty.

Rauthmann's (2024) warning against collapsing personality into any one set of self-reported traits is useful here. If even the Big Five should not be equated with the totality of personality, an Internet subtype should certainly not be treated as a complete identity.

5. The Strongest Scientific Interpretation of Personality Profiles

The person-centered literature suggests a constructive alternative to rejecting typologies entirely. Profiles can be understood as probabilistic regions in multidimensional personality space, not immutable human species.

Under this interpretation, someone may resemble one profile more strongly than another while retaining a continuous probability of membership. Two individuals assigned the same profile can differ meaningfully, and one individual's profile may change as measurement or personality changes.

This approach is closer to latent profile analysis. Profile assignment is associated with posterior probability rather than absolute essence. Person-centered results from Kerber et al. (2021), Espinoza et al. (2020), Ratchford et al. (2022), Exley et al. (2022), Preston et al. (2022), Kokkinos et al. (2023), and Lan et al. (2021) collectively support a probabilistic understanding of personality clustering.

That interpretation could also improve MBTI communication. Instead of telling a respondent, "You *are* INTJ," an evidence-sensitive formulation would be: "Your responses are currently most consistent with the INTJ preference pattern under this scoring system." The latter communicates classification without turning it into essence.

6. Implications for Organizational Management

Personality assessment can remain useful in organizations when its purpose and evidence are clearly defined. Low-risk organizational applications may include structured self-reflection, communication workshops, discussion of work preferences, coaching prompts, exploration of team differences, and career conversations in which personality information is only one input among many. In these settings, a typology can function as a shared vocabulary, and its value may lie partly in generating constructive conversation rather than in predicting performance.

The risk increases when personality categories become selection rules. If employers use subtype labels to decide who can lead, manage, sell, innovate, or collaborate, they must demonstrate criterion validity, incremental validity, fairness, and appropriate generalizability. Zárate-Torres and Correa's (2023) finding of weak MBTI–leadership relationships is particularly relevant: measurement quality does not guarantee strong leadership prediction.

Automated assessment increases both scalability and risk. Hickman et al. (2022) and Stevenor et al. (2024) show that algorithmic personality inference requires validation against appropriate criteria. High predictive accuracy inside a development sample is insufficient if the model does not generalize.

An ethical organizational rule follows from this evidence: the higher the consequence of a personality-based decision, the stronger the required evidence must be. This principle applies to human and automated decisions alike because scale does not substitute for validity, fairness, or generalizability.

7. Implications for education and career counseling

The same distinction applies in educational settings. Personality frameworks may help students discuss learning preferences, career aspirations, interpersonal behavior, or perceived strengths. Ramachandran et al. (2020), Haidar et al. (2021), and Li et al. (2024) demonstrate sustained interest in personality among healthcare learners.

However, educators should avoid deterministic statements such as "this type cannot become a surgeon," "that type should not study engineering," or "this rare subtype is naturally suited to leadership." Personality–performance relationships are probabilistic and typically modest. Meta-analytic evidence concerning broad traits supports prediction at the group level but does not justify rigid individual destiny claims.

Mammadov's (2022) large meta-analysis is informative because even robust personality–academic associations leave substantial variance unexplained. Cognitive ability, motivation, socioeconomic conditions, prior knowledge, institutional quality, health, opportunity, and life experience continue to matter.

8. Implications For Personal Users of Online MBTI Systems

Individuals do not need to abandon personality tests to use them responsibly. For personal use, an evidence-sensitive sequence is reflection → hypothesis → observation → revision rather than test result → identity → destiny. This framing treats a personality label as a revisable prompt for inquiry rather than as a fixed statement of essence or future outcome.

Users can ask whether a description resonates, examine examples from their behavior, compare feedback from others, and revisit conclusions over time. A label becomes problematic mainly when treated as immune to contradictory evidence.

The rarity component deserves particular caution. A person should not infer psychological superiority from receiving a rare type. Nor should common categories be viewed as generic or less interesting. Every individual possesses a unique configuration of traits, experiences, values, skills, motivations, relationships, and contexts that cannot be captured fully by a population category.

9. A Research Agenda for Hidden Personality Subtypes

The popularity of derivative subtype systems creates a legitimate research opportunity rather than merely a debunking exercise. First, subtype constructs should be reverse-engineered theoretically. Researchers should identify what suffixes such as AC, AH, OH, OC, or AN are claimed to measure. Are they mixtures of emotional stability, confidence, conscientiousness, attachment, affect, or interpersonal orientation? Such claims can be mapped onto established personality frameworks.

Second, incremental validity should be tested. If a subtype corresponds almost perfectly with combinations of Big Five facets, researchers should determine whether the new category predicts anything beyond those facets.

Third, latent-variable and mixture models should test whether the boundaries are empirically supported. A genuine categorical structure should produce evidence distinguishable from a continuous latent distribution artificially divided at thresholds.

Fourth, test–retest studies should assess subtype stability. If large proportions of respondents change hidden subtype within short intervals despite no major life transition, rarity estimates become difficult to interpret.

Fifth, large cross-national validation datasets are needed. García et al. (2022), Thalmayer et al. (2022), Siraji et al. (2025), and Choi et al. (2026) demonstrate methodological templates.

Sixth, behavioral outcomes should be included. Self-descriptions alone are vulnerable to shared-method variance. Informant reports, behavioral indicators, longitudinal outcomes, digital behavior, and performance criteria would strengthen validation.

Seventh, prevalence research should publish uncertainty. Rather than "INFJ-AC = 1%," a scientifically responsible output might state that an estimated

prevalence in a defined population is x% with a 95% confidence interval of y–z%, under a specified scoring algorithm.

10. A broader lesson for digital psychological culture

The debate surrounding MBTI subtypes is one instance of a broader transformation. Psychology is increasingly being translated into viral content: attachment styles, trauma types, love languages, cognitive styles, neurodivergence labels, emotional archetypes, and algorithmic identities circulate at enormous scale.

The central risk is not popularization itself. Accessible psychological language can increase curiosity and self-understanding. The problem arises when communication certainty exceeds evidence certainty.

An infographic has limited space, so it compresses assumptions. A pyramid graph communicates hierarchy instantly but may omit sampling methodology. A percentage looks objective but may hide an undefined denominator. A subtype name feels psychologically concrete even when no independent validation exists.

The appropriate solution is therefore not less public psychology but better uncertainty communication. Public-facing personality content is most defensible when the confidence of its claims is calibrated to the quality of the underlying measurement and population evidence.

CONCLUSIONS AND RECOMMENDATIONS

This qualitative literature review examined contemporary evidence relevant to MBTI-based personality typing, derivative hidden subtypes, and claims concerning personality rarity. The synthesis produces a deliberately balanced conclusion.

Recent research provides meaningful evidence concerning several psychometric properties of established MBTI dimensions, and the continued use of MBTI across counseling, education, organizations, and professional settings cannot reasonably be explained solely as scientific ignorance. At the same time, reliability and convergent validity of an established assessment do not establish the existence of every derivative subtype, prove that personality consists of sharply separated natural categories, or validate population rarity percentages.
Contemporary trait and person-centered research provides a more defensible framework. Personality characteristics can be represented dimensionally, while empirically recurring configurations can also be investigated through latent profiles and clustering. Such profiles become scientifically useful when they are reproducible, probabilistic, externally validated, and robust across samples.

The strongest conclusion concerns rarity. A claim that a hidden personality subtype represents the “top 1%,” “top 5%,” or any other population fraction is an empirical prevalence claim requiring considerably more evidence than a descriptive personality label. Construct definition, reliable measurement, structural validity, temporal stability, cross-cultural invariance, external validation, representative sampling, and uncertainty estimation are all required before such precision is warranted.

The motivating infographic should therefore be interpreted as an illustration of contemporary digital personality culture rather than as independently verified prevalence evidence. Its rarity percentages should not be treated as population estimates unless the underlying construct definitions, sampling procedures, and validation evidence are independently documented.

The most scientifically productive question is consequently not "How rare is my personality?" but "How strongly is this classification defined, measured, replicated, and validated—and what useful information does it add beyond existing personality science?" This reformulation preserves the engaging accessibility of personality typologies while aligning their interpretation with contemporary psychometric standards.

FURTHER STUDY

This research still has limitations, so further studies on the topic Beyond the 16 Types: Hidden Subtypes, Personality Rarity, and the Psychometric Limits of MBTI-Based Classification in the Digital Age—A Critical Review are needed to improve this research and expand insights for both the author and readers.

REFERENCES

Alexander, L. III, Mulfinger, E. and Oswald, F.L. (2020) 'Using Big Data and Machine Learning in Personality Measurement: Opportunities and Challenges', European Journal of Personality, 34, pp. 632–648. https://doi.org/10.1002/per.2305

Bainbridge, T.F., Ludeke, S.G. and Smillie, L.D. (2022) 'Evaluating the Big Five as an organizing framework for commonly used psychological trait scales', Journal of Personality and Social Psychology, 122(4), pp. 749–777. https://doi.org/10.1037/pspp0000395

Cemalcilar, Z. et al. (2021) 'Testing the BFI-2 in a non-WEIRD community sample', Personality and Individual Differences, 182, 111087. https://doi.org/10.1016/j.paid.2021.111087

Choi, J. et al. (2026) 'The Big Five Inventory-2 in Korea: Validation and Cross-Cultural Comparisons with the U.S. and Chinese Versions', Assessment, 33(5), pp. 866–889. https://doi.org/10.1177/10731911251357466

Choong, E.J. and Varathan, K.D. (2021) 'Predicting judging-perceiving of Myers-Briggs Type Indicator (MBTI) in online social forum', PeerJ, 9, e11382. https://doi.org/10.7717/peerj.11382

Cupani, M., Lorenzo-Seva, U. and Morán, V. (2025) 'Personality Traits in Latin America: A Cross-Cultural Study of the Big Five Factor Structure and its Relationship with Self-Reported Daily Behaviors', Journal of Personality Assessment, 107(1), pp. 89–99. https://doi.org/10.1080/00223891.2024.2353139

Dong, Y. and Dumas, D. (2020) 'Are personality measures valid for different populations? A systematic review of measurement invariance across cultures, gender, and age', Personality and Individual Differences, 160, 109956. https://doi.org/10.1016/j.paid.2020.109956

Erford, B.T. et al. (2025) 'A 25-Year Review and Psychometric Synthesis of the Myers–Briggs Type Indicator (MBTI)–Form M', Journal of Counseling & Development, 103(4), pp. 403–417. https://doi.org/10.1002/jcad.70006

Espinoza, J.A., Daljeet, K.N. and Meyer, J.P. (2020) 'Establishing the structure and replicability of personality profiles using the HEXACO-PI-R', Nature Human Behaviour, 4, pp. 713–724. https://doi.org/10.1038/s41562-020-0853-6

Exley, J. et al. (2022) 'OCEAN wealth profiles: A latent profile analysis of personality traits and financial outcomes', Personality and Individual Differences, 185, 111300. https://doi.org/10.1016/j.paid.2021.111300

Feher, A. and Vernon, P.A. (2021) 'Looking beyond the Big Five: A selective review of alternatives to the Big Five model of personality', Personality and Individual Differences, 169, 110002. https://doi.org/10.1016/j.paid.2020.110002

Gallagher, C.M. et al. (2023) 'A Short Measure of the Big Five Aspects: Development and Validation of the BFAS-40', Journal of Personality Assessment, 105(6), pp. 719–732. https://doi.org/10.1080/00223891.2022.2153690

García, L.F. et al. (2022) 'Exploring the stability of HEXACO-60 structure and the association of gender, age, and social position with personality traits across 18 countries', Journal of Personality, 90, pp. 256–276. https://doi.org/10.1111/jopy.12664

Graham, E.K. et al. (2020) 'Trajectories of Big Five Personality Traits: A Coordinated Analysis of 16 Longitudinal Samples', European Journal of Personality, 34(3), pp. 301–321. https://doi.org/10.1002/per.2259

Gupta, P. et al. (2024) 'A genome-wide investigation into the underlying genetic architecture of personality traits and overlap with psychopathology', Nature Human Behaviour, 8, pp. 2235–2249. https://doi.org/10.1038/s41562-024-01951-3

Haidar, M. et al. (2021) 'Myers-Briggs Type Indicator Personality Types of Ophthalmology Residents', Journal of Academic Ophthalmology, 13(2), pp. e158–e162. https://doi.org/10.1055/s-0041-1732346

Hall, A.N. and Matz, S.C. (2020) 'Targeting Item-level Nuances Leads to Small but Robust Improvements in Personality Prediction from Digital Footprints', European Journal of Personality, 34, pp. 873–884. https://doi.org/10.1002/per.2253

Haslam, N. et al. (2020) 'Dimensions over categories: a meta-analysis of taxometric research', Psychological Medicine, 50(9), pp. 1418–1432. https://doi.org/10.1017/S003329172000183X

Hickman, L. et al. (2022) 'Automated video interview personality assessments: Reliability, validity, and generalizability investigations', Journal of Applied Psychology, 107(8), pp. 1323–1351. https://doi.org/10.1037/apl0000695

Irwing, P. et al. (2024) 'Towards a taxonomy of personality facets', European Journal of Personality, 38(3), pp. 494–515. https://doi.org/10.1177/08902070231200919

Kang, W. et al. (2024) 'Validation of an abbreviated Big Five personality inventory at large population scale: Psychometric structure and associations with common psychiatric and neurological disorders', Comprehensive Psychiatry, 134, 152514. https://doi.org/10.1016/j.comppsych.2024.152514

Kerber, A., Roth, M. and Herzberg, P.Y. (2021) 'Personality types revisited—a literature-informed and data-driven approach to an integration of prototypical and dimensional constructs of personality description', PLOS ONE, 16(1), e0244849. https://doi.org/10.1371/journal.pone.0244849

Kokkinos, C.M., Antoniadou, N. and Voulgaridou, I. (2023) 'Personality profile differences in academic dishonesty and procrastination among Greek university students: A five factor facet-level latent profile analysis', Personality and Individual Differences, 214, 112337. https://doi.org/10.1016/j.paid.2023.112337

Kokkinos, C.M., Antoniadou, N. and Voulgaridou, I. (2024) 'Majors unleashed: unravelling students' personality profiles across academic disciplines', Current Psychology, 43, pp. 17635–17645. https://doi.org/10.1007/s12144-024-05721-2

Lan, J.B., Wong, C.S. and Zeng, G.J. (2021) 'Personality profiles for hospitality employees: Impact on job performance and satisfaction', International Journal of Hospitality Management, 98, 103018. https://doi.org/10.1016/j.ijhm.2021.103018

Li, C. et al. (2024) 'Dental students' Myers-Briggs Type Indicator personality profile in the past 50 years: Systematic review and meta-analysis', Journal of Dental Education, 88(12), pp. 1652–1665. https://doi.org/10.1002/jdd.13660

Mammadov, S. (2022) 'Big Five personality traits and academic performance: A meta-analysis', Journal of Personality, 90, pp. 222–255. https://doi.org/10.1111/jopy.12663

Marengo, D. et al. (2021) 'A meta-analysis on individual differences in primary emotional systems and Big Five personality traits', Scientific Reports, 11, 7453. https://doi.org/10.1038/s41598-021-84366-8

Mõttus, R. et al. (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), pp. 1175–1201. https://doi.org/10.1002/per.2311

Nugraha, D.Y., Widyastuti and Ridfah, A. (2025) 'Investigating the Factor Structure of the Myers-Briggs Type Indicator (MBTI): Case of Indonesian Students', Jurnal Pengukuran Psikologi dan Pendidikan Indonesia, 14(2), pp. 1–24. https://doi.org/10.15408/jp3i.v14i2.31689

Ober, T.M. et al. (2021) 'Examining the factor structure of the Big Five Inventory-2 personality domains with an adolescent sample', Psychological Assessment, 33(1), pp. 14–28. https://doi.org/10.1037/pas0000962

Otero, I., Cuadrado, D. and Martínez, A. (2020) 'Convergent and predictive validity of the Big Five factors assessed with single stimulus and quasi-ipsative questionnaires', Journal of Work and Organizational Psychology, 36(3), pp. 215–222. https://doi.org/10.5093/jwop2020a17

Pavlov, G. et al. (2021) 'Item desirability matching in forced-choice test construction', Personality and Individual Differences, 183, 111114. https://doi.org/10.1016/j.paid.2021.111114

Pires, J.G. et al. (2023) 'Preliminary validity for the Big Five Inventory-2 in Brazilian adults', Psico-USF, 28(1), pp. 91–102. https://doi.org/10.1590/1413-82712023280108

Preston, K. et al. (2022) 'Identifying family personality profiles using latent profile analysis: Relations to happiness and health', Personality and Individual Differences, 189, 111480. https://doi.org/10.1016/j.paid.2021.111480

Prinyapol, P. et al. (2023) 'Does personality categorization affect intergroup attitudes via moderation of personal values and social identity complexity mediation?', Current Issues in Personality Psychology, 12(2), pp. 120–131. https://doi.org/10.5114/cipp/156824

Ramachandran, V. et al. (2020) 'Myers-Briggs Type Indicator in Medical Education: A Narrative Review and Analysis', Health Professions Education, 6(1), pp. 31–46. https://doi.org/10.1016/j.hpe.2019.03.002

Ratchford, J.L., Ming, M.S. and Schnitker, S.A. (2022) 'Personality profiles differentially predict well-being at the facet level: A latent profile analysis of the five factor model facets', Personality and Individual Differences, 196, 111700. https://doi.org/10.1016/j.paid.2022.111700

Rauthmann, J.F. (2024) 'Personality is (so much) more than just self-reported Big Five traits', European Journal of Personality, 38(6). https://doi.org/10.1177/08902070231221853

Rhea, A.K. et al. (2022) 'An external stability audit framework to test the validity of personality prediction in AI hiring', Data Mining and Knowledge Discovery, 36(6), pp. 2153–2193. https://doi.org/10.1007/s10618-022-00861-0

Roberts, B.W. and Yoon, H.J. (2022) 'Personality Psychology', Annual Review of Psychology, 73, pp. 489–516. https://doi.org/10.1146/annurev-psych-020821-114927

Ruchensky, J.R., Edens, J.F. and Donnellan, M.B. (2025) 'Development of an Inconsistent Responding Scale for the Big Five Inventory-2', Journal of Personality Assessment, 107(3), pp. 384–391. https://doi.org/10.1080/00223891.2024.2411557

Sheppard, H. et al. (2026) 'Understanding and assessing personality across cultures: A scoping review', PLOS ONE, 21(1), e0338521. https://doi.org/10.1371/journal.pone.0338521

Siraji, M.A. et al. (2025) 'The Bangla Big Five Inventory-2: a comprehensive psychometric validation', Scientific Reports, 15, 11008. https://doi.org/10.1038/s41598-025-90264-0

Sleep, C.E., Lynam, D.R. and Miller, J.D. (2021) 'A Comparison of the Validity of Very Brief Measures of the Big Five/Five-Factor Model of Personality', Assessment, 28(3), pp. 739–758. https://doi.org/10.1177/1073191120939160

Stevenor, B.A. et al. (2024) 'Validity evidence for personality scores from algorithms trained on low-stakes verbal data and applied to high-stakes interviews', International Journal of Selection and Assessment, 32(4), pp. 544–560. https://doi.org/10.1111/ijsa.12480

Stewart, R.D. et al. (2022) 'The finer details? The predictability of life outcomes from Big Five domains, facets, and nuances', Journal of Personality, 90, pp. 167–182. https://doi.org/10.1111/jopy.12660

Strus, W. and Cieciuch, J. (2021) 'The Circumplex of Personality Metatraits and the HEXACO model: Toward refinement and integration', Journal of Personality, 89(4), pp. 803–818. https://doi.org/10.1111/jopy.12616

Temizyürek, T., Richardson, G.B. and Brown, G.R. (2024) 'Comparability of personality facets between men and women: A test of measurement invariance in IPIP-NEO facets in 49 countries', Journal of Research in Personality, 113, 104551. https://doi.org/10.1016/j.jrp.2024.104551

Thalmayer, A.G., Saucier, G. and Rotzinger, J.S. (2022) 'Absolutism, Relativism, and Universalism in Personality Traits Across Cultures: The Case of the Big Five', Journal of Cross-Cultural Psychology, 53(7–8), pp. 935–956. https://doi.org/10.1177/00220221221111813

Wetzel, E. and Frick, S. (2020) 'Comparing the validity of trait estimates from the multidimensional forced-choice format and the rating scale format', Psychological Assessment, 32(3), pp. 239–253. https://doi.org/10.1037/pas0000781

Woo, H. and Sohn, S.Y. (2022) 'A credit scoring model based on the Myers–Briggs type indicator in online peer-to-peer lending', Financial Innovation, 8, 42. https://doi.org/10.1186/s40854-022-00347-4

Wu, W. et al. (2024) 'From personality types to social labels: the impact of using MBTI on social anxiety among Chinese youth', Frontiers in Psychology, 15, 1419492. https://doi.org/10.3389/fpsyg.2024.1419492

Yang, T. and Xu, X. (2024) 'Personality profiles and political regions: A latent profile analysis approach', Current Research in Ecological and Social Psychology, 6, 100188. https://doi.org/10.1016/j.cresp.2024.100188

Yoshino, S. et al. (2022) 'A validation of the Japanese adaptation of the Big Five Inventory-2', Frontiers in Psychology, 13, 924351. https://doi.org/10.3389/fpsyg.2022.924351

Zell, E. and Lesick, T.L. (2022) 'Big five personality traits and performance: A quantitative synthesis of 50+ meta-analyses', Journal of Personality, 90(4), pp. 559–573. https://doi.org/10.1111/jopy.12683

Zárate-Torres, R. and Correa, J.C. (2023) 'How good is the Myers-Briggs Type Indicator for predicting leadership-related behaviors?', Frontiers in Psychology, 14, 940961. https://doi.org/10.3389/fpsyg.2023.940961