Personality Assessment as Spectacle:

A Debordian Critique of Institutionalized Objectivity

Paul B. Ingram, Ph.D.

Abstract

Psychological assessment is often framed as a methodological practice. Instead, it is an institutionalized epistemic regime that produces representations of persons and ideas through scores, profiles, and constructed abstractions. In doing so, assessment assumes fixed that which is dynamic and emergent. Drawing on Guy Debord's theory of the spectacle, this paper argues that personality assessment is a uniquely institutionalized component of psychological assessment. There, representations acquire and maintain ontological primacy over lived subjectivity (e.g., clinical realities) and systematic reproduction is often privileged over evolution. Through processes of commodification, ritual participation, metric fetishism, and institutional validation, personality assessment training, clinical standards, and practice culture construct a reality in which psychological life is mediated by representations. These representations are sold to the profession as objectively true all the while narrowing interpretive possibility and slowing conceptual evolution. Representations become accepted as truer than true. Situating this critique within the field's longstanding Kuhnian paradigm model, this paper argues that personality assessment traditions differ not primarily in predictive validity, but in the conceptual worlds they authorize, the standards they enforce, and the phenomena they render visible (or invisible). From a spectacle perspective, these differences reflect competing institutional logic rather than purely scientific disagreements. The paper maps key propositions of the spectacle onto assessment science, demonstrating how contemporary practices reproduce a positivistic tradition that substitutes measurement for experience and produces an inverted relation between representation and reality. Implications for how personality assessment might more critically examine its epistemic assumptions are discussed.

Keywords: *Psychological* Assessment, *Philosophy of Science, Personality Assessment*

Personality Assessment as Spectacle: A Debordian Critique of Institutionalized Objectivity

All problems in psychology are assessment problems, and there are quite a few problems (Van Dongen et al., 2026). Assessment is foundational in research (AERA, 2014). Methods are selected to establish validity and refine measurement of an ever-growing list of constructs, or their specific forms. The design considerations, statistics, data safeguards, and interpretive standards are intentional choices. Through selection, abstract concepts are operationalized through the emergent interaction of person and context (Guyon et al., 2018). Assessment equally permeates clinical practice across diagnosis, case conceptualization, and in-the-moment decision making, in psychotherapy and explicit evaluative contexts (e.g., testing). Applied use relies on various theoretical standards of evidence, process, and practice, reflecting the complication of interpreting lived experience realistically (Collins & Bunn, 2016). In this way, measurement and assessment are foundational epistemic activities which define, organize, and legitimize psychological realities (Fortunato et al., 2018).

Assessment is deeply embedded in the history and foundation of clinical psychology (Carson, 1993). As an area, it grew as a function of workforce demands and the war-industry (Gibby & Zackar, 2008). As a result, priorities in measurement embed emphases on reliability, behavioral constraint, and occupational functioning across many contemporary measures. As measurement efforts continued, celebration of technical precision was a defining feature of advancement. Rhetorical patterns over the last half century or so for this narrative tradition include, but are not limited to, emphasizing the necessity of standardization, norm and comparison group use to accommodate base rates, determinations of protocol validity, and the terms claiming various types of validity and reliability. Some initiatives have sought to formalize evidence-based psychological assessment by codifying guides on how to approach assessment practically (e.g., Bornstein, 2017; Krishnamurthy et al., 2022; Wright et al., 2022). These efforts,

as well as the problems they seek to address, identify the field's core traditions (Laudan & Carroll, 1978). In their most representative form, these efforts reflect the centrality of psychological assessment to the history and future of health service psychology.

These efforts also reveal a linguistic ecology. Jenkins (2022), for instance, notes concern that practice guidelines are "articulating consensus rather than exploring controversies" (p.23). This concern is not uncommon, complying in part with the general acceptance standards used in forensic cases (E.G., Daubert v. Merrell Dow Pharmaceuticals, 1993). Consequently, some have argued that the formalization of some areas psychological assessment may be premature for applied use, given their controversial design decisions (e.g., validity tests and dimensional models; Leonhard, 2023; Haeffel et al., 2022). The resulting binding to existing structures emerges when the institutionalized construct exceeds its scientific authority (e.g., The Diagnostic and Statistical Manual [DSM]; Cooper, 2005; Wakefield, 2016). This risk is especially evident when constructs are viewed as absolute forms rather than as a contextual artifact constructed by measurement and interpretation (Michell, 2003)

This paper focuses special attention on personality assessment as a specific domain of psychological assessment. Personality assessment serves as a focal case because it represents one of the most highly institutionalized domains of assessment within clinical psychology, characterized by extensive training traditions (Mihura et al., 2017; Ingram et al., 2020), formalized interpretive systems and ways of thinking (Wiggins, 1996; Hopwood, 2024), traditions on how data are interpreted and reported labeled (e.g., Ingram et al., 2019; Guilmette et al., 2008), and well-established professional norms (APA, 2016; Wright et al., 2017). Personality assessment is typically framed as a technical enterprise, one in which some types of instruments are labeled "objective," fostering misapplied interpretive confidence in their measurements

(Meyer & Kurtz, 2006)[1]. Personality disorder reviews, for instance, have lamented ideological adherence to models which leave clinicians blind due, in part, to intellectual silos and unproductive turf wars (Hopwood et al., 2025). Institutions organize the production of representations, which are reified through professional practice, generating epistemic inertia that narrows conceptual possibilities. The result is a form of experiential distancing, in which procedural consistency overshadows them phenomenological complexity of lived experience. Personality assessment is, therefore, a broad epistemic system which shapes how psychological reality is produced, validated, and understood (Zautra, 2025).

Personality assessment has long been described as comprising multiple interpretive traditions (Maul et al., 2016). Hopwood (2024), extending Wiggins (1996), demonstrates the paradigm pluralism (e.g., dynamic, empirical, interpersonal, etc.) found in personality assessment, with broadly similar predictive validity across distinct traditions and approaches. The perceived distinctions between these paradigms reflect differences in conceptual or methodological framing rather than substantive technical divergence. Zautra (2025), for instance, notes that assessment paradigms follow a larger philosophic distinction between their realistic, empirical, and pragmatic approaches. Outside of this methodological convergence (e.g., Laudan & Carroll, 1978), empiricism remains the dominant paradigm of personality assessment based on instrument use frequencies, a measure of public acceptance and dominant philosophy of science. Empiricism is grounded in the assumption that psychological phenomena are objectively measurable[2].

---

[1] For this reason, while any test or measure can be evaluated from this post-critical lens, I use broadband self-report instruments (often labeled "objective tests" historically and even contemporarily) due to historical factors.
[2] This epistemic stance mirrors the ontological commitments embedded in classical test theory, underscoring the way in which measurement itself is a philosophic endeavor.

Guy DeBord's Society of the Spectacle (1977) offers a framework through which the processes and practices of personality assessment can be reconsidered, conceptualizing modern social systems as spectacles — social relations mediated through representations. From this perspective, personality assessment practice may be understood as a *spectacle-producing apparatus* in which representations of people become central organizing features of professional practice, often operating in ways that remain largely unexamined. Accordingly, this paper examines personality assessment as an inversion of scientific reality, arguing that its codification of representation constrains broader philosophical engagement. This paper adopts a critical lens to examine how institutional structures, professional expectations, and shared assumptions shape interpretive practices within personality assessment, influencing how psychological reality is constructed and understood.

Personality assessment may be understood not as a single theory but as a fully institutionalized system of practices, norms, and organizations that collectively function a spectacle-producing apparatus. Rooted in a positivist tradition, the enterprise centers on the production of *spectacular objects*—tests, scales, and interpretive frameworks—that transform subjective experience into quantifiable representations. Within this tradition, interpretive foci constrain the representational scope of traits and target dimensions, reducing complex subjectivity into interpretable metrics. This reduction obscures aspects of psychological reality by translating lived experience into normative comparisons, thereby mediating clinical decision-making through representations (e.g., score interpretation) rather than through direct engagement with psychological life[3]. Thus, the spectacle represented as the scientific organization of

---

[3] There are certainly areas of assessment psychology where direct engagement is possible and more optimal (e.g., memory tasks such as making change, addressing an envelope, and so forth). No such tasks exist within personality assessment to mirror this level of observation without relying on representation (e.g., simulated situations, abstract imagery, etc.), at least so far as clinically viable tools – to my awareness.

personality assessment is the *false consciousness of personality*, existing as a representation and structured partiality of the field itself. Despite their imperfections, representations are necessary and valuable. They are incomplete and imperfect, and institutionalization exaggerates both their usefulness and their distortions, some in open defense of an institutionalized spectacle, and some blindly guided by paradigm beliefs. This view reframes the task, examining institutional organization by evaluating how the field observes, measures, and knows.

I describe the organization and influence of these spectacular objects (tests, entities, constructs) and persons as comprising "science space," a two-dimensional plane marked by protruding peaks and organized scientific ranges representing related ideas, collaborators, mentorship, and intellectual lineage that collectively comprise personality assessment's world. Bibliographic analyses, for instance, offer one way of approximating relative ontological height in science space (see Figure 2). This visual analog builds on Merton's (1968) "Mathew Effect" in which "the social structure of science provides the context for this inquiry into a complex psychosocial process that affects both the reward system and the communication system of science" (p.57). Thus, the topology of this landscape shifts over time as tectonic movements in empirical evidence, methodological innovation, and sociocultural priorities reshape the terrain, causing ideas, instruments, and scholars to rise, erode, or recede from prominence. This evolving sociocultural process reflects the science of science- how structural networks and cumulative mechanisms organize, consolidate, and transform knowledge (Fortunato et al., 2018). Ideas and instruments may become obscured as paradigms shift or grow, only to be rediscovered when new theoretical frameworks or methodological tools provide alternative vantage points on the landscape. Kopetzki et al. (2025), for instance, examined test selection from a psychoanalytic perspective, showing that the science space prioritized varies by theoretical adherence. This

pattern reflects a historically weak tradition of theory construction (Meehl, 1990). The net epistemic effect is persistent obfuscation, leaving the literature "well-nigh uninterpretable."

Navigating science space is central to navigating research careers. Movement across this terrain is structured not only by empirical discovery but by social networks, institutional positioning, and intellectual alliances (e.g., Fortunato et al., 2018; Leahey, 2007). In my observation of personality assessment science space, literature reviews are constrained largely to instruments or instrument families, rather than the constructs they measure. The research of any given instrument typically relies only on the literature and philosophies of that landscape, neatly avoiding both limitations and advancements found elsewhere. This pattern reflects a form of epistemic insularity within localized regions of science space. It results from a failure to incorporate ontological and philosophical ideas outside one's own viewpoint. Said plainly, our field lacks intersectionality and concentrates primary on cross-sectional correlations to further validate an existing measure using undergraduate students, most of whom are white, female, and interested in psychology.

Navigating science space requires the timely identification of ontological power(s), typically instructed explicitly or named as such in papers. As the relative influence of a given peak becomes more pronounced, so too does its *spectacular legitimization*. The ontological primacy[4] of that spectacle becomes justification for the representation, either directly or through increasing abstractions. To maximize ontological primacy (e.g., legitimacy), paragraphs, as well as the sentences in them, alternate between the creation and solving of problems. For example, a

---

[4] In some cases, the ontological primacy of a given test is less the rationale than the ontological need, often which relies on incorporation of some missing element from contemporary theory. For instance, while the former might reflect language indicating "a popular instrument" or "one of the most widely used measures", the later depicts issues such that no other test incorporates or measures the representation of interest, or in a manner reflective of reality. In my experience, this transition occurs approximately in the second paragraph of most papers.

test could identify a limitation in the instrument, introduce some revision as corrective (e.g., supplemental scoring, new scale, etc.), and then frame the problem as solved. The remaining limitations, most of which are rarely followed, rely largely on increasing empirical certainty about scale realism. This reliance comes at the cost of subjectivity (e.g., expansion of the construct through sub-group use)—a recuperation process (Debord, 1977). Maintenance of existing power structures is often the primary conclusion, either because of limits to one's paradigm (see Hopwood, 2024), researcher risk aversion (Fortunato et al., 2018), or a tendency not to truly reconceptualize clients (Garb, 2005; Peterson, 1996).

In such a way, research traditions manage the analogies which might give rise to more radical change in methods or measurement. Assessment outputs are interpreted historically either through some persistently evolving pattern of information gathered either from individual items, scales and indices, or configural profiles comprised of multiple informatic components drawn from scales or items. Each of these interpretive indicators are treated as more real than the nuanced psychological processes they aim to represent, reflecting the reification of constructs, the invisibility of unmeasured domains, and the simplification of experience, particularly when the construct in question holds epistemic authority within the measurement framework (see Benjamin, 2005). The result, there is a tendency towards constrained innovation and a narrowing of construct representation (e.g., Jenkins, 2022). Across personality assessment, test creation and revision has largely emphasized the retention of elements of other tests (from items to constructs). One may need to look no further than any assessment journal to compare the number of novel ideas to that of replicated constructs. In these ways, concern about maintaining and promoting existing order is often a central design concern despite evidence of methodological advancement (for instance, Nichols, 2006; *c.f.,* Weed, 2006).

The field's reliance on these spectacular products is so extensive that empirical studies frequently proceed as though measurement error is negligible, in turn violating Meehl's (1959) warning against second order validity and a lack of good criterion validity. For instance, Leonhard's (2023a) critique of symptom validity tests (SVTs) insofar as use of these measures ignores the error inherent to them, which we know of clearly (Dandachi-FitzGerald et al., 2017). Under such conditions, clinical realities are subordinated to statistical inference, increasing the risk that analytic outputs are treated as more definitive than the contexts in which they are generated. Tests take on the role of a quasi-forensic arbiter of reality based on an inferential, probabilistic architecture. This probabilistic reality is more suited to estimating likelihoods and ruling out conditions (true negatives) than to asserting definitive correspondence between scale elevations and lived experience (true positives). Constructs produce spurious interpretations as a function of over-applied generalizations between various sub-group concepts (Herring & Ingram, Accepted; Author et al., in review). Even methods are constrained in a manner that biases towards reinforcement of the instrument.

Recurring patterns across items, scales, and testing experiences are embedded within procedural rituals, including standardized instructions, interpretive conventions, and training traditions. Practices may even lack direct experimental justification (Ryan et al., 2015) but the constructs can still reinforce and legitimize the assessment enterprise and are retained accordingly. Credentialing structures grant authority while presuming objectivity, thereby supporting the continued reproduction of the system. One may look at the handling of the EPPP Stage 2 to see how difficult it is to push against engrained systems, even in light of substantive concerns about predictive validity and test bias (Saldaña et al., 2024; Sharpless, 2021). As the result of this codification of representation, momentary responses are treated as stable reality for

interpretative purposes rather than "moments in time". This pattern depicts how *false consciousness of time* (e.g., predictive stability as an assumed trait) underlies the *false consciousness of personality. There,* decisions are viewed as more valid than they are – a pattern highlighted by our decision-making accuracy (Meehl). This dynamic of falsehood, determinations affecting individual rights are framed as probabilistic language descriptions of the client's future world (Buntins et al., 2016) without true consideration of the phenomenon (see Kyriazos et al., 2024 for instance). Thus, assessment remains the scientific practice of deriving inferences from groups and applying them to individuals, despite inherently limited capacity to predict outcomes for single cases[5].

As a result of this research design dominance, clients occupy the system position of a proletariat, lacking power to shape the test or its interpretation at a systematic level. The more disenfranchised, the less likely they are to benefit from the system and, instead, "advantage accumulates" (Link & Milcarek, 1980). As one of the most vulnerable people, the client-proletariat's lived subjectivity is largely subordinated to prior patterns observed in others or assumed by theory. The language used to capture and employ these probabilistic descriptors varies greatly between clinicians (Guilmette et al., 2020; Loring et al., 2025), increasing error on either side of interpretation (e.g., the clinician producing the report, and the *clinical proletariat* impacted directly by it). While largely true of the field, work on personality functioning offers one clear example of intra-individual methods, whose findings reinforce the need and opportunity for such individually focused approaches (Hopwood et al., 2025). These structural

---

[5] While there are studies using single case design, most studies use group designs rather than intra-individual traits. There are some recent efforts to expand intra-individual design for psychological assessment needs, though most of these fall within the cognitive domain rather than personality or pathology.

conditions shape the phenomenological experience of assessment itself; however, I am unaware

of any formally incorporated intra-individual methods into published industry driven tests.

       The abstraction(s) inherent to assessment functions to distance client-proletariats from

their own lived experience, and personal truth. The result is the production of epistemic

disempowerment analogous to the *alienation of the proletariat* described by DeBord (1977). This

alienation is one reason that viable[6] normative data is critical to maintain if an instrument of high

ontological influence wishes to maintain that power (see Corrigan, 1985). Assessment

frameworks therefore embed dominant cultural norms that define what counts as typical, healthy,

or pathological, often forgoing other critical frameworks such as developmental ones (Cole &

Ingram, 2026). The pattern seen in studies is often as such – groups are combined due to poor

sampling, sub-analyses are not conducted, or they are conducted only partially. Under positivistic

assumptions (Clegg, 2016), studies often conclude constructs are valid, reinforcing the defense

of science space icons. Unrealistic classification and clear limits of any practical generalization

that would support clinical reality are endorsed, often blindly and in practices ripe with

limitations (e.g., over-reporting detection and validity determinations[7]; Leonhard, 2023a,b).

Incentive structures prioritize publication productivity over systematic examination, producing

the false consciousness of personality that is the institutionalization of ideas.

       This pattern may help explain, in part, the difficulty of disentangling processes of

disenfranchisement and distrust from the assessment procedures designed to measure them

(Dixon et al., 2023; Whaley, 2001). These issues underlie the replication crisis in assessment

---

[6] I use the term *viable* rather than *updated* because revisions do not necessarily produce viable representations. Updates in the form of new items, revised base rates, or new or refined scales or constructs may not be viable in all cases, or to the degree assumed. Even when interpretive value is added, unassessed sources of error can still obscure effectiveness, as illustrated by the example of the MMPI-3 EAT scale that follows.
[7] While Leonhard's critiques apply both to Performance and Symptom based validity (PVT and SVT, respectively), my critiques focus here on SVT given its use in personality assessment and the papers emphasis.

(Tackett et al., 2019). This recurring crisis produces statistical and conceptual artifacts that increasingly structure practice, embedding assumptions into interpretive norms and clinical decision-making (e.g., Herring & Ingram, accepted; cf. Sharf et al., 2017). For instance, Schneider et al (2026)[8] conducted a simulated over-reporting study and identify exceptionally abnormal classification accuracies (e.g., sensitivities and specificities, which are based on faulty assumptions of absolute truth – e.g., true positive and true negative are known; Dandachi-FitzGerald & Merckelbach, 2017). This reliance on simulation is unlikely to generate to clinical groups on in simulation studies (Rogers & Bender, 2020). As a result, substantial criticism of science's acceptability for practice emerges even while complimenting the technical success despite substantive limitations that impede any assumption of clinical reality (e.g., "most of the included studies use a simulation design and have been coauthored by the test authors", Puente-López et al., 2023). Likewise, Schneider et al (2026) conclude "The fact that validity cutoffs leave little room for credible emotional distress (the invalid before clinically elevated paradox) … introduces a psychometric conundrum that will need to be addressed in future research. (p.1)". My point is not to criticize these technical successes, but to draw our attention to what has at times been lost – awareness of if our efforts are to keep at bay forces which might corrupt us (see Benjamin, 2005), or to preserve a spectacle without recognizing it as such.

As in all things, distance produces perspective in science space. Altering the level of abstraction changes what can be seen and what can be reasoned, which is why thought experiments serve a legitimate scientific function. Yet consider a client whose score approximates the cut score for over-reporting invalidity and simultaneously approximates the threshold for bona fide clinical elevation. How, precisely, is such a scale to be used? Under what

---

[8] I selected this article to represent what is a large and longstanding trend across many instruments due to its recency and my role as a reviewer for the article and it's production of spectacular unfreedom.

epistemic conditions can we claim classification accuracy between 84% and 93% of cases when the very boundary separating exaggeration from pathology is abstract, probabilistic, and context-dependent- and even explicitly unknown (for instance, Schneider et al., 2026)! I cannot grant myself that level of confidence in any such hypothetical instrument.

Yet these thought experiments are repeatedly performed as if their assumptions were settled. In this way, assessment practice often functions as a "hack" (Wark, 2005)—an innovative exploitation that converts virtual statistical abstractions into institutionalized reality through its vectorial rule and encouragement of commodified production (§0168). Once codified, representations (e.g., over-reporting) acquire ontological primacy. They reorganize perception, structure professional judgment, and can displace the lived experience they were designed to approximate. Abstractions no longer merely mediate reality; they organize how reality is interpreted. Thus, "the hacker now appears in the official organs of the ruling order" (Wark, 2005, §173), albeit rarely in the form of truly novel production that characterizes creative enterprise (see Piotrowski, 2019). Such a pattern of emergence is consistent with the assessor as a spectacular figure and the instrument as a spectacular product (Debord, 1977).

Across personality assessment, concerns about predictive and conceptual validity have been documented across nearly every identifiable subgroup (Mihura & Görner, 2021; Morris et al., 2024; Liu, 2022; Ramberg et al., 2025). Sometimes due to a lack of study, this field gap has long been known (e.g., Piotrowski, 2019). These patterns are more or differently pronounced in some instruments and paradigms, yet they collectively illustrate a recurring tension between empirical anomalies and the preservation of established measurement frameworks. Measurement invariance is frequently evaluated only after publication rather than as a foundational modeling requirement (Reynolds et al., 2013). For some scholars, these findings raise questions as

consequential as the discontinuation of certain validity indicators (Leonhard & Leonhard, 2024; c.f. Merchelbach & Dandachi-FitzGerald, 2024). From a spectacle perspective, these patterns illustrate how representations persist even in the presence of systematic inconsistencies, reinforcing institutional authority while slowing conceptual revision. At the same time, emerging efforts to incorporate lived experience into construct development represent an important shift toward addressing these longstanding limitations and reducing the alienation embedded within traditional assessment frameworks (Hopwood et al., 2025; Molloy et al., 2025).

To move forward, the field must accommodate its history (e.g., Gone, 2021). Assessment is not merely a measurement enterprise but a representation-producing system, one which has contributed to the institutional authority of the profession (Benjamin, 2005). However, this spectacle-producing system carries inherent risks, as institutional reinforcement can stabilize constructs independent of their validity (**Author et al in review**). Interpretive traditions regulate epistemic authority, positioning clients in structurally subordinate roles. Meanwhile, the assessment process delegates clinicians as ritual participants in ontological reproduction. The result is that impactful lived experience is often treated as pathology, or otherwise unassessed (Dixon et al., 2023), despite clear calls for more qualitative considerations given this ontological tendency (Miller, 2003). In this way, the delegation of design and interpretive authority produce flawed normative frameworks. Before pursuing reform, the field must therefore clarify which elements of assessment reflect necessary representations and which reflect the inertia of institutional reproduction. This distinction ultimately determines whether assessment evolves as a scientific enterprise or merely stabilizes its own representations.

Below I offer some ways to advance personality assessment research and practice. These recommendations are situated within the framework that assessments produce representations

with ontological influence, displacing lived realities (DeBord, 1977). This displacement occurs when group realities are inferred as individual truth, despite limited evidence of single case predictive accuracy. The task is not to abandon assessment, but to situate it more explicitly within its epistemic limits and social conditions. Assessment psychology has a lot of problems that we can choose to fix.

A greater appreciation of systematic processes is needed to move the human sciences forward (Gamsakhurdia, 2026), including assessment and personality science. Positionality is critical and Reflexivity statements are one mechanism for expressing these systematic influences in one's work, assuming they are not used as proxies for methodological rigor or devolve into performative "check boxes" (Singh, 2025). Rather, statements and efforts to understand positionality must be taken seriously, addressing how instrument selection, construct commitments, training background, and institutional context shape the design, execution, interpretation, and publication of a study. Objections to expanding reflexivity often rest on the concern that it undermines assumptions of empirical realism and the image of the mechanical, neutral researcher (Savolainen et al., 2023)— threatening the spectacle itself (Debord, 1977). In contrast, positionality merely makes available for analysis the science-time forces that have already shaped the scientific worldview and its claims to objectivity. From this vein, open peer review should be considered— an effort to make transparent the spectacle's production (e.g., publication). Admittingly, there is no easy definition of open peer review that could be adopted to address concerns of institutionalization on objectivity (Ross-Hellauer, 2017; Tennant & Ross-Hellauer, 2020). Nevertheless, assessment science should pursue a model better suited to its epistemic commitments and economic realities— one that renders visible the scientific reasoning of reviewers who are as ontological forces, even if that reasoning is substantively absent.

Moreover, designs should incorporate multiple elements of conceptual or methodological need (Piotrowski, 2019). Bibliometric evaluation should be more common, and more comprehensive so as to include the broader range of science space influences. Designs should likewise more explicitly articulate the limits of group-derived inference when applied to single cases, clarifying where probabilistic classification ends and lived subjectivity begins. Greater incorporation of intra-individual and context-sensitive designs may help temper the reification of traits by situating psychological phenomena within their developmental and situational realities (e.g., Oblak et al., 2026). For instance, how fully has the field capitalized on opportunities to refine measurement accuracy at the level of item construction itself? Might normative procedures account for the impact of word choice and stem construction on valence and arousal (e.g., Bradley & Lang, 1999). Even subtle shifts in evaluative tone may alter affective activation and, consequently, response patterns like social desirability.

Validation efforts should more routinely interrogate the cultural and developmental assumptions embedded within normative frameworks, particularly before claims of invariance or generalizability are institutionalized. Artificial Intelligence, for instance, provides a unique and timely opportunity to continuously accommodate local lived experience (data) and emerging science (norms) in item construction itself (see Herring & Ingram, Accepted), advancing and hacking *itemmetrics* (see Johnson, 2005; Wiggins & Goldberg, 1965). Encouraging cross-instrument and construct-level synthesis, rather than remaining within instrument-bound literatures, may reduce the epistemic insularity that stabilizes peaks in science space. More should be done to encourage novelty, discourage ritualized reproduction of existing constructs, and reconsider the incentives that reward spectacular production over conceptual revision. Embracing these, or other, approaches may reveal our collective *unfreedom* - our blind adherence

to constructs, ideas, or practices beyond their evidence. These reflective self-explorations are as important for researchers as they are for editorial teams, if we wish to produce open science (Sharpe, 2024).

References

Author et al (in review)

American Educational Research Association, American Psychological Association, & National

      Council on Measurement in Education. (2014). *Standards for educational and*

      psychological testing. American Educational Research Association.

American Psychological Association. (2016). Standards for accreditation for health service

      psychology. American Psychological Association.

      https://www.apa.org/ed/accreditation/about/policies/standards-of-accreditation.pdf

Benjamin Jr, L. T. (2005). A history of clinical psychology as a profession in America (and a

      glimpse at its future). *Annu. Rev. Clin. Psychol.*, *1*, 1-30.

Bornstein, R. F. (2017). Evidence-Based Psychological Assessment. *Journal of Personality*

      Assessment, *99*(4), 435–445. https://doi.org/10.1080/00223891.2016.1236343

Bradley, M.M., & Lang, P.J. (1999). Affective norms for English words (ANEW): Instruction

      manual and affective ratings. Technical Report C-1, The Center for Research in

      Psychophysiology, University of Florida

Buntins, M., Buntins, K., & Eggert, F. (2016). Psychological tests from a (fuzzy-) logical point

      of view. *Quality & Quantity*, *50*(6), 2395-2416.

Clegg, J.W. (2016). Reconsidering philosophy of science pedagogy in psychology: An evaluation

      of methods texts. *Journal of Theoretical and Philosophical Psychology, 36(4), 199.*

Cooper, R. (2004). What is wrong with the DSM?. *History of Psychiatry*, *15*(1), 5-25.

Colligan, R. C. (1985). History and development of the MMPI. *Psychiatric Annals*, *15*(9), 524-

      533.

Collins, A., & Bunn, G. (2016). The shackles of practice: History of psychology, research

assessment, and the curriculum. In *Centrality of history for theory construction in psychology* (pp. 91-109). Cham: Springer International Publishing.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

Dixon, J. N., Caddell, T. M., Alexander, A. A., Burchett, D., Anderson, J. L., Marek, R. J., & Glassmire, D. M. (2023). Adapting assessment processes to consider cultural mistrust in forensic practices: An example with the MMPI instruments. *Law and Human Behavior*, *47*(1), 292.

Dandachi-FitzGerald, B., Merckelbach, H., & Ponds, R. W. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *Journal of Clinical and Experimental Neuropsychology*, *39*(3), 257-264.

Krishnamurthy, R., Hass, G. A., Natoli, A. P., Smith, B. L., Arbisi, P. A., & Gottfried, E. D. (2022). Professional practice guidelines for personality assessment. *Journal of Personality Assessment*, *104*(1), 1-16.

Kyriazos, T., & Poga, M. (2024). Exploring fuzzy logic as an alternative approach in psychological scoring. *The Open Psychology Journal*, *17*(1).

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., ... & Barabási, A. L. (2018). Science of science. *Science*, *359*(6379), eaao0185.

Gamsakhurdia, V. L. (2026). Dialogical and systemic anthro-psychology: Toward a culturally sensitive transdisciplinary human science. *Theory & Psychology*, 09593543261418607.

Garb, H. N. (2005). Clinical judgment and decision making. *Annu. Rev. Clin. Psychol. 2005*, *1*, 67-89.

Gibby, R. E., & Zickar, M. J. (2008). A history of the early days of personality testing in American industry: An obsession with adjustment. *History of psychology*, *11*(3), 164.

Gone, J. P. (2021). Decolonization as methodological innovation in counseling psychology: Method, power, and process in reclaiming American Indian therapeutic traditions. *Journal of Counseling Psychology*, *68*(3), 259.

Guilmette, T. J., Hagan, L. D., & Giuliano, A. J. (2008). Assigning qualitative descriptions to test scores in neuropsychology: Forensic implications. *The Clinical Neuropsychologist*, *22*(1), 122-139.

Guilmette, T. J., Sweet, J. J., Hebben, N., Koltai, D., Mahone, E. M., Spiegler, B. J., ... & Conference Participants. (2020). American Academy of Clinical Neuropsychology consensus conference statement on uniform labeling of performance test scores. *The Clinical Neuropsychologist*, *34*(3), 437-453.

Guyon, H., Kop, J., Juhel, J., & Falissard, B. (2018). Measurement, ontology, and epistemology: Psychology needs pragmatism-realism. *Theory & Psychology,* 28(2).

Haeffel, G. J., Jeronimus, B. F., Kaiser, B. N., Weaver, L. J., Soyster, P. D., Fisher, A. J., ... & Lu, W. (2022). Folk classification and factor rotations: Whales, sharks, and the problems with the Hierarchical Taxonomy of Psychopathology (HiTOP). *Clinical Psychological Science*, *10*(2), 259-278.

Herring, T.T., & Ingram, P.B (Accepted). Distinction without a difference: Do symptom validity tests differentially measure symptom presentations? *Current Treatment Options in Psychiatry.*

Herring & Ingram (Accepted). Ethics of the Use and Integration of Artificial Intelligence (AI) into Psychological Assessment. *Professional Psychology: Research and Practice.*

Hopwood, C. J., Kaurin, A., & Pincus, A. L. (2026). Revitalizing personality disorder

research. *Journal of Psychopathology and Clinical Science*, *135*(1), 1.

Johnson, J.A. (2004). The Impact of Item Characteristics on Item and Scale Validity.
Multivariate Behavioral Research, 39(2), 273-302.

Kopetzki, A., Blais, M. A., de Saeger, H., Finn, S. E., Lewis, K. C., & Hopwood, C. J. (2025).
Test selection in psychoanalytically informed assessment. *Psychoanalytic Psychology*.

Laudan, L., & Carroll, P. T. (1978). Progress and its problems: Toward a theory of scientific
growth.

Leahey, E. (2007). Not by productivity alone: How visibility and specialization contribute to
academic earnings. *American Sociological Review, 72*, 533–561.

Leonhard, C. (2023a). Review of statistical and methodological issues in the forensic prediction
of malingering from validity tests: Part I: Statistical issues. *Neuropsychology
Review*, *33*(3), 581-603.

Leonhard, C. (2023b). Review of statistical and methodological issues in the forensic prediction
of malingering from validity tests: Part II—Methodological issues. *Neuropsychology
Review*, *33*(3), 604-623.

Link, B., & Milcarek, B. (1980). Selection factors in the dispensation of therapy: The Matthew
effect in the allocation of mental health resources. *Journal of Health and Social Behavior*,
279-290.

Leonhard, C., & Leonhard, C. (2024). Through Smoke and Mirrors: Excluding Malingering
Expert Testimony Under the Daubert Standard. *Ga. L. Rev.*, *59*, 1.

Loring, D. W., Hermann, B. P., Meador, K. J., Lah, J. J., Goldstein, F. C., & Bilder, R. M. (2025).
Would a rose by any other name smell as sweet? Complexity, context, and consequences
of neuropsychology performance labels. *The Clinical Neuropsychologist*, 1-9.

Lui, P. P. (2022). Whose evidence? Enhancing cultural competency and humility in personality

    assessment: Commentary on Krishnamurthy et al.(2022). *Journal of Personality*

    Assessment, *104*(1), 19-22.

Maul, A., Irribarra, D.T., & Wilson, M. (2016). On the philosophical foundations of

    psychological measurement. *Measurement,* 79, 311-320

Meehl, P. E. (1959). Some ruminations on the validation of clinical procedures. *Canadian*

    Journal of Psychology, 13(2), 102–128.

Merckelbach, H., & Dandachi-FitzGerald, B. (2024). The (small) world of validity test research

    according to Leonhard and Leonhard (2024): A critique. *Psychological Injury and*

    Law, *17*(4), 439-449.

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of

    science are considered. *Science*, *159*(3810), 56-63.

Meyer, G., & Kurtz, J.E. (2006). Advancing personality assessment terminology: Time to retire

    "objective" and "projective" as personality test descriptors. *Journal of Pesonality*

    Assessment

Michell, J. (2003). The quantitative imperative: Positivism, naïve realism, and the place of

    qualitative methods in psychology. Theory & Psychology, 13, 5-31.

Mihura, J. L., & Görner, K. J. (2021). Assessment of Personality and Psychopathology: Key

    Challenges and Systemic Improvements Needed. In J. L. Mihura (Ed.), *The Oxford*

    handbook of personality and psychopathology assessment (2nd ed., pp. 918–946). Oxford

    University Press. https://doi.org/10.1093/oxfordhb/9780190092689.013.40

    If you want, I can also generate:

Mihura, J. L., Meyer, G. J., Dumitrascu, N., & Bombel, G. (2013). The validity of individual

Rorschach variables: Systematic reviews and meta-analyses of the comprehensive system. *Psychological bulletin*, *139*(3), 548.

Molloy, N., Kilcoyne, I., Belcher, H., & Wykes, T. (2025). Exploring the involvement of people with lived experience of mental disorders in co-developing outcome measures: a systematic review. *The Lancet Psychiatry*, *12*(2), 140-152.

Morris, C. S., Shepherd, E. A., & Ingram, P. B. (2024). Investigating the validity of the MMPI-3 Eating Concerns (EAT) scale across men and women in a university sample. *Journal of Personality assessment*, *106*(1), 17-26.

Nichols, D. S. (2006). The trials of separating bath water from baby: A review and critique of the MMPI–2 Restructured Clinical Scales. *Journal of Personality Assessment*, *87*(2), 121-138.

Oblak, A., Vrbnjak, M., Holnthaner, A., Kovačič, N., Kastelic, M. P., Marvin Derganc, T., ... & Škodlar, B. (2025). From self-narration to a worldview: a phenomenological, narratological, and linguistic case study of a patient with a complex clinical picture of bipolar disorder. *Frontiers in Psychiatry*, *16*, 1648141.

Piotrowski, C. (2019). Contemporary research emphasis in personality assessment: A bibliometric analysis mapping investigatory domain (2009-2018). *Journal of Projective Psychology & Mental Health*, *26*(2), 98-104.

Puente-López, E., Pina, D., López-Nicolás, R., Iguacel, I., & Arce, R. (2023). The Inventory of Problems–29 (IOP-29): A systematic review and bivariate diagnostic test accuracy meta-analysis. *Psychological Assessment*, *35*(4), 339.

Ramberg, C. E., Menton, W. H., Arbisi, P. A., Polusny, M. A., & Marquardt, C. A. (2025).

Validity of the Minnesota Multiphasic Personality Inventory–3 Eating Concerns (EAT) Scale: A study with National Guard veterans and their romantic partners. *Psychological Services*.

Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research*, *6*, 588.

Ryan, J. J., Swopes-Willhite, N., Franklin, C., & Kreiner, D. S. (2015). WAIS-IV Administration Errors: Effects of Altered Response Requirements on Symbol Search and Violation of Standard Surface-Variety Patterns on Block Design. *Applied Neuropsychology: Adult*, *22*(1), 42-45.

Saldaña, S., Callahan, J. L., & Cox, R. J. (2024). The Examination for the Professional Practice of Psychology: An examination of construct validity. *Training and Education in Professional Psychology*, *18*(1), 42.

Savolainen, J., Casey, P. J., McBrayer, J. P., & Schwerdtle, P. N. (2023). Positionality and its problems: Questioning the value of reflexivity statements in research. *Perspectives on Psychological Science*, *18*(6), 1331-1338.

Sharpe, D. (2024). Editor bias and transparency in psychology's open science era. *American Psychologist*, *79*(7), 883.

Singh, A., Dechant, M. J., Patel, D., Soubutts, E., Barbareschi, G., Ayobi, A., & Newhouse, N. (2025, April). Exploring positionality in HCI: Perspectives, trends, and challenges. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-18).

Sharpless, B. A. (2021). Pass rates on the Examination for Professional Practice in Psychology

(EPPP) according to demographic variables: A partial replication. *Training and Education in Professional Psychology*, *15*(1), 18.

Tackett, J. L., Brandes, C. M., King, K. M., & Markon, K. E. (2019). Psychology's replication crisis and clinical psychological science. *Annual review of clinical psychology*, *15*(1), 579-604.

Tennant, J. P., & Ross-Hellauer, T. (2020). The limitations to our understanding of peer review. *Research integrity and peer review*, *5*(1), 6.

Wakefield, J. C. (2016). Diagnostic issues and controversies in DSM-5: return of the false positives problem. *Annual review of clinical psychology*, *12*(1), 105-132.

Wark, M. (2004). *A hacker manifesto*. Harvard University Press.

Weed, N. C. (2006). Syndromal complexity, paradigm shifts, and the future of validation research: Comments on Nichols and Rogers, Sewell, Harrison, and Jordan. *Journal of Personality Assessment*, *87*(2), 217-222.

Whaley, A. L. (2001). Cultural mistrust: An important psychological construct for diagnosis and treatment of African Americans. *Professional Psychology: Research and Practice*, *32*(6), 555.

Wiggins, J.S. & Goldberg, L.R. (1965). Interrelationships among MMPI item characteristics. *Educational and Psychological Measurement*, 25, 381-397.

Wright, A. J., Pade, H., Gottfried, E. D., Arbisi, P. A., McCord, D. M., & Wygant, D. B. (2022). Evidence-based clinical psychological assessment (EBCPA): Review of current state of the literature and best practices. *Professional Psychology: Research and Practice*, *53*(4), 372.

Wright, C. V., Beattie, S. G., Galper, D. I., Church, A. S., Bufka, L. F., Brabender, V. M., &

Smith, B. L. (2017). Assessment practices of professional psychologists: Results of a national survey. *Professional psychology: research and practice*, *48*(2), 73.

Zautra, N. (2025). Psychiatry's New Validity Crisis: The Problem of Disparate Validation. *Philosophy of Science*, *92*(3), 646-665.

Figure 1. Sample Topographic Visualization of Science Space

Note. The "heights" of the so-called science space mountains represent values extracted from Table 1 in Piotrowski (2019). The figure includes selected constructs of interest as indexed in that table; however, it does not arrange those constructs into organized ranges, nor does it depict less frequently measured ideas, individual scholars, institutional influences, field organization dynamics, or other structuring forces within science space. Rather, it serves as a simplified visual analog to illustrate how differential emphasis can be rendered topographically when introducing the science space metaphor. This image was created using ChatGPT.