Please use the following citation when referencing this work:

Dombrowski, S. C., & McGill, R. J. (2024). Clinical assessment in school psychology: Impervious to scientific reform? *Canadian Journal of School Psychology*. Advance online publication. https://doi.org/10.1177/08295735231224052

Clinical Assessment in School Psychology: Impervious to Scientific Reform?

Stefan C. Dombrowski[1]

Ryan J. McGill[2]

[1]Department of Graduate Education, Leadership, and Counseling, Rider University

[2] Department of School Psychology and Counselor Education, William & Mary

Author Note

Stefan C. Dombrowski https://orcid.org/0000-0002-8057-3751

Ryan J. McGill https://orcid.org/0000-0002-5138-0694

Author Contribution Statement: All authors were responsible for the design, conceptualization, and writing of the manuscript.

Informed Consent/Patient Consent: No patients were used in the writing of this manuscript so informed consent is not required.

Conflict of Interest Disclosure: None.

There are no data to report.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

All procedures were conducted in accordance with the ethical standards in publishing.

Please address all correspondence concerning this article to Stefan C. Dombrowski, Rider University, Department of Graduate Education, Leadership, and Counseling, 2083 Lawrenceville Road, Lawrenceville, NJ 08648; sdombrowski@rider.edu

Clinical Assessment in School Psychology: Impervious to Scientific Reform?

Throughout the past century, the promise of school psychology to progress as a scientific discipline has been discussed at various futures conferences and in scholarly articles (e.g., Conoley et al., 2020). While some areas of school psychology have experienced progress (e.g., prevention science, school-based mental health), other sectors have either stagnated or barely progressed at all from their original conceptualizations. An area of school psychology where science has progressed slowly by comparison is clinical assessment, specifically cognitive assessment in general and the identification of specific learning disability (SLD) in particular. Given the latter remains a prominent focus of training and practice, this lack of progress is concerning given that problematic assessment practices have existed in this space since the inception of the profession.

Pertaining to ability assessment, while there have been some gains since the late 1980s delimiting the use of selected low-value interpretive practices (e.g., subtest pattern analysis, scatter analysis), broader cognitive profile analysis remains alive, if not, thriving in contemporary practice. For instance, cognitive profiles are used as a form of evidence for a new class of ill-specified conditions (e.g., trauma, adverse childhood experiences) despite nascent evidence to the contrary (Farmer et al., 2022). In totality, interest in cognitive profile analysis persists even though the available research evidence to date contraindicates their propagation in any form (McGill et al., 2018; Watkins, 2000).

Regarding SLD identification, although the field seems poised to cast aside the IQ-achievement (IAD) discrepancy, this move has generated several new approaches to identification (e.g., patterns of strengths and weakness [PSW] analysis, response to intervention [RTI], and alternative dual deficit low achievement models subsequently adopted by the DSM-

5). Given this degree of variability with respect to SLD identification it is important to consider that assessment and associated activities (e.g., report writing) remain the predominant activity of psychologists in the schools and non-evidence-based practice is endemic for test interpretation and diagnostic decision-making for SLD (Benson et al., 2020; Dombrowski et al., 2022).[1] Accordingly, the purpose of this brief essay is to address the question of whether assessment in school psychology can progress as a true scientific discipline by highlighting some of the ecosystemic factors that influence assessment training and practice.

Stagnant Progress and Lack of Consequential Validity

For decades, the field of assessment research has produced an array of blueprints, articles, and perspectives from conferences that offer a pathway for evidence-based assessment (EBA). The field has also borne witness to waves of novel approaches to assessment that augured to overcome prior limitations. For example, there are new or newly revised instruments and assessment algorithms and programs that claim practice improvement and linkage to theory (e.g., Flanagan et al., 2022). Virtually all of these technologies are disseminated as research-based to some degree.

However, the promoted methods have done little to meaningfully advance the field's ability to address the "big problems" in school psychology (Shapiro, 2000). In spite of long-standing calls for a paradigm shift (e.g., Reschly & Ysseldyke, 2002), scientific progress has been exceptionally slow and it sometimes seems as if the field is merely spinning in circles giving the appearance of momentum but without much demonstrable progress. It is well

[1] It is recognized that clinical assessment encompasses numerous low-inference applications not discussed here such as behavioral assessment, direct observation, and curriculum-based assessment. Their omission is not an indication of lack of importance. Rather, the focus of this article is on what is regarded as the most salient examples of an evidence to practice gap.

understood that stakeholders do not set out on a path leading to disciplinary inertia; rather, it is a consequence of numerous factors. In what follows, some possible explanations leading to the present state of clinical assessment practice are outlined.

Theoretical and Methodological Incoherence

Clinical assessment in school psychology is theoretically complex, if not incoherent, residing at the intersection of many disciplines including clinical psychology, special education, neuropsychology, cognitive science, neuroscience, measurement, pediatrics, biopsychology, and basic research in reading, writing, and mathematics. This leads to an overwhelming literature base where researchers are frequently unaware of each other's exploratory silos. This context also permits a milieu where researchers can extrapolate literature from one field (e.g., cognitive neuroscience and cognitive psychology) to make a theoretical justification for an assessment practice (e.g., PSW) with subsequent correlational support to buttress speculative claims offered in scholarly forums. Researchers are then able to justifiably attach the moniker "research-based" to their assessment methods which is true in part; however, this ad hoc approach to theory development leads to conditions where it is difficult to envision how theories and approaches to test interpretation and use can be falsified. Since it is rare for any method of investigation in assessment research to utilize strong inference to address pertinent questions, this may yield a series of conflicting findings using different methods with no connectivity to a systematic program of research (Kranzler et al., 2016).

Long ago, Meehl (1990) warned about theory building in the social sciences using correlational research and the reification of what he termed the "crud factor" or the ability to obtain non-zero correlations for virtually anything in sufficiently large data sets as a basis for deriving psychologically meaningless constructs and then subsequently using those constructs as

a reference marker in future validation studies. This genre of construct validation is more illusory than grounded in evidence though it tends to be impenetrable to dismantling once it becomes engrained in the literature through a seemingly connected but fragmented evidentiary chain featuring weak, but high-probability predictions (e.g., a score on this cognitive variable will significantly predict a related score on an achievement test that is moderately to highly correlated). By contrast, strong theories make point predictions (i.e., my knowledge of this test score will lead to a correct diagnosis or treatment) which are amenable to falsification through risky empirical tests.

Relatedly, it is common for a new test to be developed with an accompanying technical manual reporting an abundance of psychometric information supporting its use. Subsequent independent evaluations of the instrument's theoretical structure (i.e., factor structure) yield a divergent series of possible rival interpretive structures for the test. After a decade or more, the test is revised without regard for rival models presented in the independent literature and therefore receives limited adjudication of the instrument's true measurement model. Although connectivity and replication is highly valued in psychological science (Shaw et al., 2019) it may be difficult to ascertain when scale development (and theory generation) proceeds in this fashion.

Take, for instance, the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V; Wechsler, 2014). A five-factor structure, cohering with Cattell-Horn Carroll Theory (CHC; Schneider & McGrew, 2018), has been posited by the test publisher. However, subsequent, numerous independent research has found predominantly an alternative four-factor structure (see Dombrowski et al., 2021c). What is the explanation for these disparate findings? Will the test publisher, in future renditions, evaluate these findings or at least explore the rationale for them? For instance, the disparity could be related to nuances *between* the methods employed for

investigating structural validity (e.g., exploratory factor analysis [EFA] vs. confirmatory factor analysis [CFA]), each requiring different underlying assumptions. It could also be related to differences *within* the method employed (e.g., EFA) which may contain distinct iteration algorithms that can be triggered by problematic indicators (Dombrowski et al., 2019)[2]. In this methodological morass it is not surprising that some scholars (e.g., Decker et al., 2020) have suggested—incorrectly—that certain estimation techniques may be biased, ostensibly on the basis that they do not return the publisher's preferred results for an instrument. Further complicating matters is the emergence of a new class of visualization techniques (e.g., network analysis) as a vehicle of structural validity which may or may not cohere with factor analysis results.

In totality, the field should not view the conclusions from a singular structural validity study, regardless of the methodology employed or independence of researcher, as the *Dei verbum*. Rival theoretical/interpretive models emerge in the literature. Some of these models may replicate in simulation studies (Dombrowski et al., 2021b) and yet others seem to be the mere product of capitalizing on the 'statistical production function' (i.e., chance; c.f., Meehl, 1978) with any given sample. Thus, the results from factor analysis, like any method that falls within the general linear model (e.g., multiple regression), requires substantive replication before an instrument or algorithm is adopted widely. However, this may be problematic on a practical front as practitioners are encouraged to adopt newly revised instruments within a year of their publication despite likely not having a firm understanding of what the test actually measures. Also, the few researchers with expertise to examine an instrument's structure have competing

[2] Beaujean (2016) and Dombrowski et al. (2021a) have also raised concern that some preferred models may be an artifact of the degrees of freedom afforded to researchers using methods (e.g., CFA) that require substantive closed-door user input.

attention for their time (e.g., teaching; administrative) so many instruments remain bereft of independent evaluation leaving no choice to the practitioner but to rely upon the information presented in the technical manual as the basis for interpretation for several years or more.

Influence of Commercial and Legal Interests

Adding to the complexity, assessment in school psychology in the United States is often governed by various state and federal regulations (e.g., IDEA, 2004). For instance, school psychologists are presented with several choices for SLD identification including PSW, IAD, RTI, and other *research-based* alternatives. This creates confusion for practitioners and stakeholders attempting to advance EBA as it has led to a degree of diagnostic chaos in the field (Schroeder et al., 2017). In addition, the codification of various procedures in IDEA potentially offers greater leverage for commercial gain and procurement of grant funding as entrepreneurs both inside and outside of the formal world of academia are incentivized to respond to evolving market interests. When a practice is legally codified then it tends to gain a degree of social validity. As an example, look to the various research groups that have benefitted professionally since the codification of RTI in the federal guidelines in the United States. This should not be construed as a value judgment on the matter; it is a simple observation. Nevertheless, Lilienfeld (2017) noted that the incentive structures in academia "make our myopic focus on grant funding counterproductive to scientific progress" (p. 633). While Lilienfeld's perspective may be unduly harsh, it merits consideration.

Further, there appears to be an infinite number of possibilities to propagate a practice as 'evidence-based' so long as a user can demonstrate the method has some modicum of a research basis undergirding it. However, evidential quality for these justifications vary considerably. In some cases, they are predicated upon nothing more than invoking the term itself as a sort of

magical incantation. Additionally, the incoherent knowledge base described previously makes it all too easy for self-appointed gurus to cherry pick from a variety of data to support or dismiss virtually any claim creating an illusion of knowledge for users (Leng & Leng, 2020). This problem is amplified in a digital world where dissemination of scientific information has been democratized and there are few if any barriers to entry to becoming a practice influencer. As an example, look to the number of digital assessment experts that emerged almost immediately following the outbreak of the pandemic regardless that this area of assessment research was and continues to be under-developed (Farmer et al. 2021a).

With respect to SLD identification, it is important to understand that, absent a gold-standard consensus operational definition, the field has been chasing an empirical phantom since the inception of the condition. Given the almost limitless definitional boundaries, a discreet definition and SLD diagnostic approach becomes ever more elusive emboldening academic and professional granfalloons to create a narrative that their assessment approach works. This ecosystem is antithetical to aspirational Mertonian norms whereby scholars remain sufficiently disinterested in their findings permitting both negative and positive evidence to inform change regardless of what the evidence indicates. Relatedly, this intractability is accentuated by potential conflicts of interest that remain largely unaddressed in school psychology (Beaujean et al., 2022). Accordingly, Boccaccini et al. (2017) warn “we would be wise to consider the possible role of researcher allegiance when we read any individual study and look for findings that generalize across studies to guide clinical practice” (p. 336).

Promotion and Dissemination of Assessment Practices

In this complex ecosystem, practitioners attempt to discern the evidence base; however, it is likely that the widespread dissemination of procedures in workshops and at conferences by

adherents of various assessment practices overshadows the reach of an article in a scholarly journal no matter the article's evidentiary quality. The wide availability of information promoting a particular practice as research-based may contribute to a perspective that all perspectives are valid since there is often research for and against a practice (i.e., the jury is still out). Willingham (2012) called out this approach to scientific inquiry labelling it the simple egalitarian fallacy[3]. Whereas an understanding of multiple perspectives is an important trait of school psychologists, it is not one that serves the field well when attempting to vet the scientific literature. It is indeed appropriate to be open to new ideas and multiple perspectives; yet, this vantage should be counterbalanced by healthy skepticism (Lilienfeld et al., 2012) and a review of evidential quality, particularly when the vast majority of data is proprietary and open access to it is often circumscribed.

State of Assessment and Methodological Training of School Psychologists

Finally, an area that deserves further exploration is the state of assessment training in school psychology. Lockwood and Farmer (2020) surveyed the tests commonly used in school psychology training programs and concluded that while there has been some progress, the use of problematic measures (e.g., projective measures) remains. More recently, a content analysis of textbooks commonly used for cognitive assessment training (Farmer et al., 2021b) found that the majority of sources continued to promote low-value interpretive practices such as profile analysis. Thus, it is not surprising that surveys (e.g. Kranzler et al., 2020) find that a disproportionate number of practitioners continue to engage in these practices raising the question as to whether they are impervious to dismantling.

[3] The simple egalitarian fallacy is defined as giving multiple perspectives due consideration even though they should not be afforded such respect based upon the evidence available.

As utilizing theory *and* research is a core imperative undergirding the EBA movement, it is reasonable to interrogate the degree to which practitioners are adequately trained in the methodological foundations of measurement, providing them with the skills necessary to navigate the complexities of the assessment literature. In a survey of specialist-level programs included in the National Association of School Psychologists (NASP) database, McGill and Wilson (2017) found that only ~10% of training programs listed a standalone measurement course in their program of studies and that the entire research sequence was limited to a single semester-long combined course in methods/statistics in most programs. Although the principles of measurement may be emphasized in other areas of the curriculum (i.e., assessment courses), it is important to question whether placing only secondary emphasis on the foundations of measurement is sufficient in an assessment landscape that is growing ever more complex. Put simply, assessment training in school psychology appears to be perpetuating low value practices, not inoculating future practitioners against them.

Conclusion

Given the transdisciplinary influences on school psychology along with its requirement to comport with federal and state law in the U.S., scientific progress in the area of cognitive assessment and SLD identification is unlikely to mature as a science for the near future without substantive reform. Extrapolation of research from one domain to the realm of assessment is common in school psychology where test authors and creators of interpretive and diagnostic systems make theoretical justification for their claims and support those claims with either multiple regression methodology that is prone to find significant associations given large samples or factor analysis that can uncover a preferred structure depending upon methodological choice and assumptions utilized. Greater replication and reproduction of such research is

necessary prior to adopting a theory or practice. Federal and state law in the U.S. permits low value practices that practitioners readily adopt since they are legal. School psychology does not have a regulatory body like the Health Products and Food Branch (HPFB) in Canada to oversee practices. All of these issues impinge on scientific progress in cognitive assessment and SLD identification.

More extensive training in critical thinking and scientific decision-making as well as measurement (e.g., factor analysis, diagnostic validity, and receiver operating curve analysis) may serve to inoculate school psychologists against low value practices. However, it is important to recognize that in psychology, problematic theories and ideas do not abate when they are falsified; instead, they tend to fade away when a sufficient number of people gradually lose interest (Meehl, 1978). Although it may be tempting to place some of the blame for entrenched, low value practices and theories on the entrepreneurial interests that pervade our field, it is instead posited that commercial interests are creating a recursive loop that not only promotes practice, but also responds to the needs of the field. Therefore, it will behoove the field to look inward and recognize that change is likely to emerge from stakeholders and practitioners within the discipline. Nonetheless, until U.S. federal law (e.g., IDEA) distances itself from non-evidence-based practice it is likely that scientific progress will transpire slowly. Fortunately, Canada is not beholden to omnibus special education law, which serves as a platform for practitioner's justification for the use of low value practice, so its academic institutions and practice agencies (e.g., school districts) may be better poised to serve as a beacon of scientific light leading the way toward scientific progress in cognitive assessment and SLD identification.

References

Beaujean, A. A. (2016). Reproducing the Wechsler Intelligence Scale for Children-Fifth Edition: Factor model results. *Journal of Psychoeducational Assessment, 34*(4), 404-408. https://doi.org/10.1177/0734282916642679

Beaujean, A. A., McGill, R. J., & Dombrowski, S. C. (2022). Financial conflicts of interest in school psychology: A continuing problem. Contemporary School Psychology. Advance online publication. https://doi.org/10.1007/s40688-022-00435-3

Benson, N. F., Maki, K. E., Floyd, R. G., Eckert, T. L., Kranzler, J. H., & Fefer, S. A. (2020). A national survey of school psychologists' practices in identifying specific learning disabilities. *School Psychology, 35*(2), 146-157. https://doi.org/10.1037/spq0000344

Boccaccini, M. T., Marcus, D., & Murrie, D. C. (2017). Allegiance effects in clinical psychology research and practice. In S. O. Lilienfeld & I. D. Waldman (Eds.), *Psychological Science Under Scrutiny* (pp. 323-339). John Wiley. https://doi.org/10.1002/9781119095910.ch16

Conoley, J. C., Powers, K., & Gutkin, T. B. (2020). How is school psychology doing: Why hasn't school psychology realized its promise? *School Psychology, 35*(6), 367-374. https://doi.org/10.1037/spq0000404

Decker, S. L., Bridges, R. M., Luedke, J. C., & Eason, M. J. (2021). Dimensional evaluation of cognitive measures: Methodological confounds and theoretical concerns. *Journal of Psychoeducational Assessment, 39*(1), 3-27. https://doi.org/10.1177/0734282920940879

Dombrowski, S. C., McGill, R. J., Canivez, G. L., Watkins, M. W., & A. A. Beaujean (2021a). Factor analysis and variance partitioning in intelligence research: Clarifying misconceptions. *Journal of Psychoeducational Assessment, 39*(1), 28-38. https://doi.org/10.1177/0734282920961952

Dombrowski, S. C., McGill, R. J., & Morgan, G. B. (2021b). Monte Carlo modeling of contemporary intelligence test (IQ) factor structure: Implications for IQ assessment, interpretation, and theory. *Assessment, 28*(3), 977-993. https://doi.org/10.1177/1073191119869828

Dombrowski, S. C., McGill, R. J., Watkins, M. W., Canivez, G. L., Pritchard, A. E., & Jacobson, L. A. (2021c). Will the real theoretical structure of the WISC-V please stand up? Implications for clinical interpretation. *Contemporary School Psychology, 26*, 492-503. https://doi.org/10.1007/s40688-021-00365-6

Dombrowski, S. C., McGill, R. J., Farmer, R. L., Kranzler, J. H., & Canivez, G. L. (2022). Beyond the rhetoric of evidence-based assessment: A framework for critical thinking in clinical practice. *School Psychology Review, 51*(6), 771-784. https://doi.org/10.1080/2372966X.2021.1960126

Farmer, R. L., McGill, R. J., Dombrowski, S. C., Benson, N. F., Smith-Kellen, S., Lockwood, A. B., Powell, S., Pynn, C., & Stinnett, T. A. (2021a). Conducting psychoeducational assessments during the COVID-19 crisis: The danger of good intentions. *Contemporary School Psychology, 25*(1), 27-32. https://doi.org/10.1007/s40688-020-00293-x

Farmer, R. L., McGill, R. J., Dombrowski, S. C., & Canivez, G. L, (2021b). Why questionable assessment practices remain popular in school psychology: Instructional materials as pedagogic vehicles. *Canadian Journal of School Psychology, 36*(2), 98-114. https://doi.org/10.1177/0829573520978111

Farmer, R. L., Zaheer, I., & Schulte, M. (2022). Disentangling low-value practices from pseudoscience in health service psychology. *Philosophical Psychology*. Advance online publication. https://doi.org/10.1080/09515089.2022.2144193

Flanagan, D. P., Alfonso, V. C., Zinkiewicz, G. J., Ortiz, S. O., & Dynda, A. M. (2022). An integrative theoretical framework for cognitive test interpretation in school neuropsychological evaluations. In D. C. Miller, D. E. Maricle, C. L. Bedord, & J. A. Gettman (Eds.). *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based interventions* (2nd ed., pp. 87-118.). Wiley.

Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016). Cross-Battery Assessment pattern of strengths and weaknesses approach to the identification of specific learning disorders: Evidence-based practice or pseudoscience? *International Journal of School & Educational Psychology, 4*(3), 146-157. https://doi.org/10.1080/21683603.2016.1192855

Kranzler, J. H., Maki, K. E., Benson, N. F., Eckert, T. L., Floyd, R. G., & Fefer, S. A. (2020). How Do School Psychologists Interpret Intelligence Tests for the Identification of Specific Learning Disabilities? Contemporary School Psychology, 24(4), 445-456. https://doi.org/10.1007/s40688-020-00274-0

Leng, G., & Leng, R. I. (2020). *The matter of facts: Skepticism, persuasion, and evidence in science*. MIT Press.

Lilienfeld, S. O. (2017). Psychology's replication crisis and the grant culture: Righting the ship. *Perspectives on Psychological Science, 12*(4), 660-664. https://doi.org/10.1177/1745691616687745

Lilienfeld, S. O., Ammirati, R., & David, M. (2012). Distinguishing science from pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology, 50(*1), 7-36. https://doi.org/10.1016/j.jsp.2011.09.006

Lockwood, A. B., & Farmer, R. L. (2020). The cognitive assessment course: Two decades later. *Psychology in the Schools, 57*(2), 265-283. https://doi.org/10.1002/pits.22298

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*(4), 806-834. https://doi.org/10.1037/0022-006X.46.4.806

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 66*(1), 195-244. https://doi.org/10.2466/PR0.66.1.195-244

McGill, R. J., & Wilson, K. M. (2017, February). *Specialist-level training in statistics, measurement, and methodology in school psychology: Implications for evidence-based practice* [Poster session]. Meeting of the Trainers of School Psychologists, San Antonio, TX.

McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology, 71*, 108-121. https://doi.org/10.1016/j.jsp.2018.10.007

Reschly, D. J., & Ysseldyke, J. E. (2002). Paradigm shift: The past is not the future. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 3-20). National Association of School Psychologists.

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73-163). Guilford.

Schroeder, M., Drefs, M. A., & Cormier, D. C. (2017). The messiness of LD identification: Contributions of diagnostic criteria and clinical judgment. *Canadian Psychology, 58*(3), 218-227. https://doi.org/10.1037/cap0000115

Shapiro, E. S. (2000). School psychology from an instructional perspective: Solving big, not little problems. *School Psychology Review, 29*(4), 560-572. https://doi.org/10.1080/02796015.2000.12086043

Shaw, S. R., S. D'Intino, J., & Lysenko, E. (2019). Registered reports, replication, and the Canadian Journal of School Psychology: Improving the evidence in evidence-based school psychology. *Canadian Journal of School Psychology, 34*(3), 175-187. https://doi.org/10.1177/0829573519843027

Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly, 15*(4), 465-479. https://doi.org/10.1037/h0088802

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). Pearson.

Willingham, D. T. (2012). *When can you trust the experts? How to tell good science from bad in education*. Wiley.