

**Why Questionable Assessment Practices Remain Popular in School Psychology: Instructional Materials
as Pedagogic Vehicles**

Ryan L. Farmer¹, Ryan J. McGill², Stefan C. Dombrowski³, & Gary L. Canivez⁴

1 Oklahoma State University, Stillwater, OK; <http://orcid.org/0000-0003-1409-7555>

2 William and Mary, Williamsburg, VA; <https://orcid.org/0000-0002-5138-0694>

3 Rider University, Lawrenceville, NJ; <https://orcid.org/0000-0002-8057-3751>

4 Eastern Illinois Univeristy, Charleston, IL; <https://orcid.org/0000-0002-5347-6534>

This is a preprint. The peer-reviewed version of this article is in press at the ***Canadian Journal of School Psychology*** and will be published as part of a 2021 special issue on De-implementation in School Psychology. This document will be updated as soon as a DOI is available.

Correspondence concerning this manuscript should be addressed to Ryan L. Farmer, Oklahoma State University School of Teaching, Learning, and Educational Sciences, Stillwater, OK 74078. E-mail: r.farmer@okstate.edu.

COI Statement: R. L. Farmer contributed to one textbook (Kranzler & Floyd, 2020) and R. J. McGill served as a reviewer for that textbook and received a free copy of that textbook by the test publisher for those efforts. G. L. Canivez contributed to a second textbook (Canivez, 2013) reviewed in this manuscript.

Note: Special thanks to Ashley Hale for her assistance with formatting and copyediting.

Preliminary results were previously presented at the 2019 meeting of the Trainers of School Psychologists, Atlanta, GA,

Abstract

Surveys reveal that many school psychologists continue to employ cognitive profile analysis despite the long-standing history of negative research results from this class of practice. This begets the question: why do questionable assessment practices persist in school psychology? To provide insight on this dilemma, this article presents the results of a content analyses of available interpretive resources in the clinical assessment literature that may shed insight on this issue. Although previous reviews have evaluated the content of individual assessment courses, this is the first systematic review of pedagogical resources frequently adopted in reading lists by course instructors. The interpretive guidance offered across tests within these texts was largely homogenous emphasizing the primary interpretation of subscale scores, de-emphasizing interpretation of global composites (i.e., FSIQ), and advocating for the use of some variant of profile analysis to interpret scores and score profiles. Implications for advancing evidence-based assessment in school psychology training and guarding against unwarranted unsupported claims in clinical assessment is discussed.

Why Questionable Assessment Practices Remain Popular in School Psychology: Instructional materials as Pedagogic Vehicles

Despite their ubiquity (see Kranzler et al., 2016), intelligence (IQ) test use and interpretation is controversial (e.g., Beaujean & Benson, 2019; Fiorello et al., 2007; McGill et al., 2018), with some questioning whether IQ tests should be used at all (Gresham & Witt, 1997; see Fletcher & Miciak, 2017 for a more nuanced perspective). Even a casual inspection of the literature reveals that there are numerous interpretive approaches that are available to aid clinicians as they navigate the complex array of primary and ancillary scores produced by modern IQ tests. For instance, there are numerous profile analysis-based interpretive systems (e.g., cross-battery assessment, levels-of-analysis approach) that, despite their popularity (Kranzler et al., 2020), are the subject of extensive discourse due to concerns over the “evidence-base” that has been extended by proponents to support their use.

A Brief Review of the Evidentiary Status of IQ Interpretation Strategies

It is beyond the scope of this article to discuss, at length, the evidentiary status of the various IQ interpretation strategies. However, this section serves as an overview of comprehensive reviews for interested readers and summarizes supporting evidence that certain interpretive strategies should be regarded as *low-value practices*. In de-implementation research, low-value practices are those that (a) lack evidence of effectiveness or are not efficacious, (b) are less effective or efficacious than another practice with the same function, (c) cause harm, or (d) are no longer necessary (e.g., McKay et al., 2018). With regard to (a) and (b), effective and efficacious are defined in the context of intelligence testing as strategies with diagnostic or treatment utility or incremental validity for predicting outcomes.

As such, we would conclude based on the available evidence that Stratum I (i.e., subtest-level; see Watkins, 2000; 2003) and Stratum II (composite-level; see McGill et al., 2018 and Watkins, 2000) profile analysis strategies are low-value practices because they (a) are not adequately supported by compelling empirical evidence and (b) alternative approaches such as low-inference assessment (e.g.,

CBM, functional assessment) better serve the intended function of these approaches to test interpretation (e.g., Fletcher & Miciak, 2017). Additionally, the practice of ignoring or giving little interpretive weight to Stratum III scores (i.e., global IQ scores) in favor of primary interpretation of Stratum II dimensions (e.g., see Kaufman & Lichtenberger, 2005); or, of disregarding the Stratum III scores when constituent parts are significantly different (see McGill, 2017; Schneider & Roman, 2017) lack evidence of incremental validity, and would also be classified as low-value practices. In a recent review evaluating psychometric and conceptual concerns regarding IQ test interpretive practices, McGill et al. (2018) concluded that primary interpretation of subscale scores may be misguided as independent structural validity studies indicate that many of these scores are not adequately located by popular IQ tests and, even when located, often lack sufficient unique reliable variance for confident clinical interpretation. Furthermore, although modern IQ tests are multidimensional, results from independent factor-analytic studies indicate that most of the reliable variance at all levels of IQ tests is explained primarily by general intelligence and not by Stratum II constructs. These results replicate similar shortcomings noted in previous reviews (e.g., Watkins, 2000). Put simply, the numerous shortcomings identified in the body of literature weaken most, if not all, of the foundational assumptions undergirding the use of profile analysis techniques.

Consistent with Floyd and Kranzler (2019), McGill et al. (2018), and Watkins (2000), this conclusion does not entail that research programs investigating underlying theory (e.g. Cattell-Horn-Carroll [CHC]; see Schneider & McGrew, 2018), approaches (e.g., patterns of strengths and weaknesses), or aptitude by treatment interactions (ATIs) are unimportant research lines or pseudoscientific as a matter of course; but, that their ability to advance clinical *practice* is currently limited. In addition to the interpretation of composites and heuristics regarding when to interpret global IQs, many tests produce numerous ancillary scores that are not supported by theory (Beaujean & Benson, 2019). The lack of a theoretical basis and psychometric adequacy for interpretation makes interpretation of ancillary scores

contraindicated. In sum, appeals to theory do not obviate the need to ensure that scores produced by IQ tests have a baseline-level of appropriate psychometric support. Prevailing ethical guidelines and codes (e.g., American Educational Research Association et al., [2014]) make this point clear. Unfortunately, the psychometric information furnished in some test technical manuals does not even meet *de minimus* standards for adequate reporting, making the task of determining whether a test or individual scores should be used for high-stakes decision-making futile (McGill et al., 2020).

State of Practice & Training

Despite the long-standing psychometric and conceptual issues associated with interpretation of subscale scores in general, and the use of profile analysis methods in particular, surveys reveal that these interpretive practices remain in use. For example, Sotelo-Dynega and Dixon (2014) surveyed 323 practicing school psychologists and found that about half followed a levels-of-analysis approach and a quarter applied the cross-battery assessment framework. Additionally, about half (45%) of their participants disregarded the global IQ due to significant scatter and 1% reported never interpreting the global IQ score at all. However, more than half (56%) reported that they interpreted composite scores all of the time. More recently, Kranzler et al. (2020) surveyed 1,317 practicing school psychologists regarding their use of IQ test interpretation strategies as part of specific learning disability assessments. They found that most clinicians interpreted profiles of subtest scores (~69%) and/or index scores (~64%), or generally apply a levels-of-analysis approach (~29%). While the majority (~80%) of participants reported interpreting the global IQ, more than half (~62%) reported not interpreting the global IQ in the presence of scatter. These data may suggest a reversal of interpretation patterns from those Sotelo-Dynega and Dixon (2014) observed; alternatively, these data may reflect the differences in sampling methods and methodology employed by the two studies. Specifically, the differences observed may be due to Sotelo-Dynega and Dixon's (2014) focus on general interpretation practices whereas Kranzler et al. (2020) limited their focus to cases where specific learning disability was the primary

classification of concern. Perhaps there is a difference in how school psychologists interpret IQ tests when they use them to identify intellectual or developmental disabilities versus when they use them to identify specific learning disabilities. Regardless, both studies suggest that the interpretation of profiles, the interpretation of composite scores, and the disregard of global IQ in the presence of scatter are common practice.

While assessment practices come from a variety of sources, we focus on training experiences given that trainers maintain influence in that domain. Cook and colleagues (2009) surveyed 2,607 psychotherapists in the United States and Canada to identify variables that may influence clinical practices and the adoption of evidence-based practices. The most influential variables the authors identified were clinicians' mentors, books, graduate training, and discussions with peers. These findings may generalize to school psychology, with two-thirds of clinicians reporting they used strategies learned during their graduate coursework and from test technical manuals (Sotelo-Dynega & Dixon, 2014). As graduate training and textbook exposure appear to have a significant impact on IQ test interpretation strategies, it is important to consider how students are taught to interpret such tests and the books assigned to facilitate and guide that coursework and future self-guided professional development once they enter the field.

Fortunately, two studies have described how IQ testing is taught in school psychology programs (Lockwood & Farmer, 2019; Miller et al., 2020). These studies investigated not only the textbooks that are commonly used, but also the type of the interpretive strategies that are typically taught within training programs. The results of these studies suggested that significant emphasis is placed upon IQ test cognitive profile analysis, and that the majority of sources cited to support this practice were not peer reviewed (e.g., McGill et al., 2018). The Lockwood and Farmer (2019) study surveyed 127 graduate trainers responsible for teaching coursework on IQ testing in school psychology programs. Results indicated more than two-thirds of trainers teach students to interpret Stratum II composites in isolation

and about two-thirds teach students to compare those composites. This is consistent with additional data suggesting that approximately 69% of trainers teach some form of patterns of strengths and weaknesses analysis and 39% of trainers teach the “Intelligent Testing” framework first introduced by Kaufman (1979) over 40 years ago. In addition, Lockwood and Farmer (2019) found that approximately one-third of trainers teach students to interpret subtest scores and to compare subtest scores. These data seem to support the premise that low-value interpretive strategies continue to be taught in graduate coursework for IQ testing.

In addition, the understanding of which textbooks clinicians used in their graduate courses may further illuminate and inform why IQ test interpretation practices with little scientific support continue to be popular in practice. Miller et al. (2020) collected syllabi from 90 graduate trainers regarding their programs’ IQ testing course. Various versions of Sattler’s *Assessment of Children: Cognitive Foundations*; Flanagan et al., *Contemporary Intellectual Assessment: Theories, tests, and issues*; Kranzler and Floyd’s *Assessing Intelligence in Children and Adolescents: A Practical Guide*; Schrank et al., *Essentials of WJ-IV Cognitive Abilities Assessment*; and Flanagan and Alfonso’s *Essentials of WISC-V Assessment* were the most frequently required textbooks. Most of the frequently used textbooks on IQ testing provided a detailed explication of stratum II and stratum III analyses. The practices commonly described within these textbooks involve the following stepwise analyses: (a) interpreting Stratum II scores and profiles, (b) interpreting Stratum I scores and profiles, (c) disregarding Stratum III scores in the presence of scatter, and (d) interpreting ancillary scores (collectively referred to as low-value practices) despite a substantial amount of counterfactual evidence in some cases (e.g., McGill et al., 2018).

Purpose of the Present Study

Because there is a dearth of supportive, peer-reviewed research for (a) interpreting Stratum II scores and profiles, (b) interpreting Stratum I scores and profiles, (c) disregarding Stratum III scores in the presence of scatter, and (d) interpreting ancillary scores (collectively referred to as low-value

practices) and substantial amount of counterfactual evidence available in some cases (e.g., McGill et al., 2018), and data that suggests these strategies continue to be explicitly taught in graduate programs (Lockwood & Farmer, 2019), we hypothesized that non peer-reviewed sources (i.e., textbooks and test manuals; henceforth, instructional materials) overwhelmingly recommend these interpretive methods be used in clinical practice. Historically, recommendations of low-value practices have been included in some test manuals (e.g., Wechsler, 2014) and some frequently used textbooks (e.g., *Essentials of Cross-Battery Assessment*) (Miller et al., 2020) center on these practices. However, it is unclear to what extent the guidance available in instructional materials for the most commonly administered IQ tests (see Benson et al., 2019; Sotelo-Dynega & Dixon, 2014) align with available peer-reviewed research evidence pursuant to these matters (Lilienfeld et al., 2006).

The purpose of this study was to evaluate available instructional materials to identify the interpretive procedures recommended for clinicians within and between contemporary IQ tests. Although the major goal was to classify themes related to the interpretive guidance featured most prominently across available resources, isolating and amplifying where particular aspects of interpretive practices may have evolved was also included. The present investigation yields information for trainers and assessment scholars when considering which resources to adopt for future intelligence assessment courses.

Methods

The present study employed a content analysis approach (Hsieh & Shannon, 2005) to code instructional materials that were selected for inclusion. Target resources included prominent books, chapters, test technical manuals, and third-party guidebooks (i.e., the *Essentials* series) for the five most-commonly used commercial ability measures at child-age (see Benson et al., 2019; Sotelo-Dynega & Dixon, 2014). These measures included the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V), Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG), Kaufman Assessment Battery

for Children-Second Edition (KABC-II), Differential Ability Scales-Second Edition (DAS-II), and Stanford-Binet Intelligence Scales-Fifth Edition (SB5). To ensure adequate saturation, internet and library searches were conducted in the Fall of 2019 using each tests' acronym and for general intellectual assessment textbooks; one textbook was updated in 2020. Additionally, the reference lists from chapters for individual tests were also screened to locate other potential sources for inclusion in the present review. In order to be included in the present study, the resource had to systematically describe clinical interpretation procedures (i.e., step-by-step) for use with a particular instrument or across instruments. In total, 34 instructional materials were selected for inclusion. Chapters and general frameworks were identified for inclusion, even when those materials were present in the same resource (e.g., Sattler, 2018). A systematic framework was then developed to code the instructional materials (see Table 1).

Results

Of the 34 interpretive resources identified providing step-by-step interpretive guidelines, seven focused on the WISC-V; five focused on the WJ IV COG; five focused on the KABC-II; four focused on the DAS-II; six focused on the SB5; and seven provided general guidance on cognitive test interpretation. Sattler (2018) was included for the WISC-V, WJ IV COG, DAS-II, SB5, and general guidance reviews as separate chapters for these instruments were provided that included unique guidance by each test consistent with the framework for interpretation discussed throughout the textbook. Details of coding for each included resource are organized by test or general guidance and are presented in Tables 2 through 7. For brevity, we will focus on instructional materials that provide general guidance.

Seven different resources were identified that provided general guidance rather than test-specific guidance. Of those, three (Canivez, 2013; Glutting et al., 2003; Kranzler & Floyd, 2020) encouraged clinicians to focus their interpretation mostly on Stratum III scores. Three (Flanagan et al., 2013; Flanagan et al., 2008; Hale & Fiorello, 2004) took the opposite position, discouraging interpretation of Stratum III scores, and instead encouraged clinicians to focus their interpretation on

Stratum II scores. One (Miller & Maricle, 2019) did not address Stratum III scores at all and encouraged a levels of analysis approach with primary emphasis at Stratum II. Three of these resources (Flanagan et al., 2013; Hale & Fiorello, 2004; Miller, 2013) invoked the variability hypothesis at Stratum II—Miller also expressed concerns for ancillary composites. With regard to ancillary composites, most authors either did not address them at all or suggested they should be interpreted with caution. Flanagan et al. (2008) encouraged their interpretation as part of a step-by-step, levels of analysis approach, though retreated to a more cautious position in future resources (Flanagan et al., 2013). Miller (2013), however, suggested that ancillary scores should be interpreted as part of a levels of analysis approach, and that the interpretive value of ancillary scores were superior to other scores. Most (excluding Hale & Fiorello, 2003; Miller, 2013) discouraged the development of inferences from individual items, whereas all who mentioned test session behavior encouraged the generation of inferences.

Discussion

The present examination identified several themes in instructional materials that seem to support many of the low-value IQ test interpretation practices employed by clinicians (Kranzler et al., 2020; Sotelo-Dynega & Dixon, 2014) which remain a staple in many training programs (Lockwood & Farmer, 2019). First, the majority of the instructional materials surveyed recommended that Stratum II scores should be the primary focal point for clinical interpretation and that interpretation of omnibus, full scale scores was often de-emphasized as a result, despite that the majority of IQ test variance partitioning research clearly shows that general intelligence explains the vast majority of variance in most of these indicators (e.g., Dombrowski et al., 2020) and Stratum II dimensions almost never contain sufficient portions of unique variance for confident interpretation (Canivez & Youngstrom, 2019). Only three resources specifically recommend against Stratum II score interpretation and profile analysis methods more generally. The homogeneity of interpretive strategies presented across the instructional materials is disconcerting but may well predict trends in practice and instruction.

Despite evidence contradicting the use of Stratum I interpretation being available since the 1990s (see Watkins, 2000) and former advocates recommending against interpretation at this level (e.g., Kaufman et al., 2016) entirely, interpretive guidance regarding Stratum I varied across instructional materials with a narrow majority (56%) suggesting such subordinate scores are interpretable in a levels of analysis approach. Others recommended Stratum I be interpreted with caution while one (Schrank et al., 2016) suggested that their interpretive value was superior to other scores. Similar results were obtained regarding the variability hypothesis. Whereas the vast majority of resources encouraged examiners to forgo interpretation of composite scores in the presence of significant test scatter, more recent resources (e.g., Drozdick et al., 2018; Flanagan & Alfonso, 2017) noted shortcomings associated with this practice. Of note, Drozdick et al. (2018) deserve special mention as they reversed course and wrote that this specific practice was likely un-categorically unsupported based on recent research (e.g., McGill, 2016). In total, these findings suggest that *some* self-correction has occurred in the last 20 years with respect to subtest analysis and the variability hypothesis. However, the results of the present study illustrate that despite this positive momentum, the vast majority of instructional materials continue to recommend low-value practices.

In sum, results from the present study suggest that the information contained within popular textbooks and manuals does not always align with the assessment practices recommended in the peer-reviewed literature, with some instructional materials promoting a greater amount of non-empirically supported practices than others. Unfortunately, lack of scientific self-correction in academic textbooks is not uncommon; particularly, in texts that focus on the status and functioning of human intellectual abilities and their measurement (Warne et al., 2018). This is consistent with the insight of Meehl (1978) who noted that popular clinical practices are passed down from generation to generation of practitioners through clinical lore and become almost immune to self-correction. Indeed, several of the instructional materials reviewed in this study have been mainstays in cognitive assessment coursework

for decades (e.g., Sattler's and Kaufman's textbooks are prominent now [Lockwood & Farmer, 2019; Miller et al., 2020], and were prominent in both the 1980s [Oakland & Zimmerman, 1986] and 1990s [e.g., Alfonso et al., 2000]) with little change in the overall interpretive recommendations offered through those resources across the decades despite accumulating contrary evidence (e.g., Watkins, 2000). The aforementioned minor change may then be due to a collective loss of interest in a specific set of practices rather than an accumulation of the evidence-base (Meehl, 1978). Instead, low-value practices seem to be reified and recycled (McGill et al., 2018).

Trainers should consider whether instructional materials have been responsive to the research literature when adopting course materials and are encouraged to give greater consideration to *empirical* resources (i.e., peer-reviewed articles) where countering evidence is presented and discussed. Continued reliance on conventional training resources will likely perpetuate low-value practices and a preference for assessment practices that have been empirically questioned and, in some cases, discredited in the literature (Truscott et al., 2004; Youngstrom, 2013). Given this predicament, in the spirit of Meichenbaum and Lilienfeld (2018), we present a provisional list of potential warning signs for hype in the clinical assessment literature as well as an annotated bibliography of seminal resources on these matters (https://osf.io/cs9jz/?view_only=7b59c13393c1440a954f0c8871ff5ab9) as a safeguard against the adoption and perpetuation of contraindicated practices.

Limitations & Future Research

The following limitations should be considered. First, selection of instructional materials may have been incomplete or biased in some way (e.g., overlooking texts from other fields, including clinical psychology or I/O psychology). While this may be true and future research should address any gaps in inclusion criteria, there was significant overlap with the instructional materials identified and those identified as commonly required or recommended by instructors (cf. Miller et al., 2020). Second, a premise of this study was that several common interpretive practices (see Sotelo-Dynega & Dixon, 2014;

Kranzler et al., 2020) lack adequate empirical support, and therefore qualify as low-value interpretive practices. While this premise is well supported (see Cohen, 1959; McGill et al., 2018; Watkins, 2000, 2003), different evidentiary criteria may result in the inclusion and exclusion of various practices. For example, it may be that some researchers may argue that simulation studies or factor analysis are poor evidence for utility and validity (e.g., McGrew, 2018). Third, guidance by referral concern are not separated (e.g., guidance for the identification of intellectual disability versus guidance for the identification of specific learning disability). While doing so may lead to greater clarity, instructional material either contained or did not contain guidance on low-value practices and so it was decided not to approach the task in this manner. Finally, while there is evidence that the identified instructional materials are used by trainers (Miller et al., 2020) and that trainers are teaching about low-value interpretive practices (Lockwood & Farmer, 2019), it is not clear whether trainers are providing this content because of specific state or district policy and also including discussion of counterfactual evidence and caution. Future researchers may be interested in exploring the context in which these materials and strategies are taught in graduate coursework.

Conclusion

Instructional materials such as assessment-specific textbooks and technical manuals potentially serve as “pedagogic vehicles” (Kuhn, 1967, p. 137) for low-value practices such as cognitive profile analysis, the interpretation of ancillary scores, or other scores or comparisons. Given the potential influence of textbooks and graduate coursework on the long-term professional behavior of clinicians (Cook et al., 2009; Sotelo-Dynega & Dixon, 2014), trainers who teach cognitive assessment should be aware of this risk and wary about how such content is presented to students. However, if history provides any indication, it is likely that the influence of non-empirical resources will continue to countervail the influence of emerging evidence-based assessment movements in scientific psychology (e.g., Youngstrom, 2013). Lilienfeld et al. (2017) contend that the incorporation of scientific thinking into

graduate training may help to reduce the scientist-practitioner gap, and thus trainers have a responsibility to foster scientific thinking. Accordingly, we encourage trainers to explicitly teach students to detect inflated claims in the assessment literature (see Lilienfeld et al., 2012; Michenbaum & Lilienfeld, 2018) and to select material that incorporates the peer-reviewed literature, including peer-reviewed articles themselves, to inoculate against hype in the clinical assessment literature.

References

* Reviewed materials

* Alfonso, V. C., & Flanagan, D. P. (2006). Best practices in the use of the Stanford-Binet Intelligence Scales, (SB5) with preschoolers. *Psychoeducational assessment of preschool children*, 267-295.

Alfonso, V. C., LaRocca, R., Oakland, T. D., & Spanakos, A. (2000). The course on individual cognitive assessment. *School Psychology Review*, 29(1), 52-64.

American Educational Research Association, American Psychological Association, & National Council on Measurement on Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Beaujean, A. A., & Benson, N. F. (2019). Theoretically-consistent cognitive ability test development and score interpretation. *Contemporary School Psychology*, 23, 126-137.

<https://doi.org/10.1007/s40688-018-0182-1>

Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 National Survey. *Journal of School Psychology*, 72, 29–48.

<https://doi.org/10.1016/j.jsp.2018.12.004>

Benson, N. F., Maki, K. E., Floyd, R. G., Eckert, T. L., Kranzler, J. H., & Fefer, S. A. (2019). A national survey of school psychologists' practices in identifying specific learning disabilities. *School Psychology*, 35(2), 146-157. <https://doi.org/10.1037/spq0000344>

* Canivez, G. L. (2013). *Psychometric versus actuarial interpretation of intelligence and related aptitude batteries*. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwann (Eds.), *Oxford library of psychology. The Oxford handbook of child psychological assessment* (p. 84–112). Oxford University Press.

Canivez, G. L., & Youngstrom, E. A. (2019). Challenges to the Cattell-Horn-Carroll Theory: Empirical, clinical, and policy implications. *Applied Measurement in Education, 32*(3), 232–248.

<https://doi.org/10.1080/08957347.2019.1619562>

Cohen, J. (1959). The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *Journal of Consulting Psychology, 23*(4), 285–299. <https://doi.org/10.1037/h0043898>

Cook, J. M., Schnurr, P. P., Biyanova, T., & Coyne, J. C. (2009). Apples don't fall far from the tree: Influences on psychotherapists' adoption and sustained use of new therapies. *Psychiatric Services, 60*(5), 671-676.

* Decker, S. L., Hale, J. B., & Flanagan, D. P. (2013). Professional practice issues in the assessment of cognitive functioning for educational applications. *Psychology in the Schools, 50*(3), 300-313. <https://doi.org/10.1002/pits.21675>

Dombrowski, S. C., McGill, R. J., Canivez, G. C., Watkins, M. W., & Beaujean, A. A. (2020). Factor analysis and variance partitioning in intelligence test research: Clarifying misconceptions. *Journal of Psychoeducational Assessment*. Advanced online publication. <https://doi.org/10.1177/0734282920961952>

* Drozdick, L. W., Singer, J. K., Lichtenberger, E. O., Kaufman, J. C., Kaufman, A. S., & Kaufman, N. L. (2018). The Kaufman Assessment Battery for Children-Second Edition and KABC-II Normative Update. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 333-359). Guilford.

* Dumont, R., Willis, J. O., & Elliott, C. D. (2008). *Essentials of DAS-II assessment* (Vol. 72). John Wiley & Sons.

Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2007). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant?. *Applied Neuropsychology, 14*(1), 2-12.

- * Flanagan, D. P., & Alfonso, V. C.. (2016). *WJ IV Clinical Use and Interpretation: Scientist-practitioner Perspectives*. Academic Press.
- * Flanagan, D. P., & Alfonso, V. C. (2017). *Essentials of WISC-V assessment*. Hoboken, NJ: Wiley
- * Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (Vol. 84). John Wiley & Sons.
- * Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Dynda, A. M. (2008). Best practices in cognitive assessment. *Best practices in school psychology V, 3*, 633-660.
- Fletcher, J. M., & Miciak, J. (2017). Comprehensive cognitive assessments are not necessary for the identification and treatment of learning disabilities. *Archives of Clinical Neuropsychology, 32*(1), 2-7. <https://doi.org/10.1093/arclin/acw103>
- Floyd, R. G., & Kranzler, J. H. (2019). Remediating student learning problems: Aptitude by treatment interaction vs. skill by treatment interaction. In M. K. Burns (Ed.), *Introduction to school psychology: Controversies and current practice* (pp. 413-434). New York, NY: Oxford University Press.
- * Glutting, J. J., Watkins, M. W., & Youngstrom, E. A. (2003). *Multifactorial and cross-battery ability assessments: Are they worth the effort?* In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement* (p. 343–374). The Guilford Press.
- Gresham, F. M., & Witt, J. C. (1997). Utility of intelligence tests for treatment planning, classification, and placement decisions: Recent empirical findings and future directions. *School Psychology Quarterly, 12*(3), 249–267. <https://doi.org/10.1037/h0088961>
- * Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. Guilford Press.
- Hsieh, H., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277-1288. <https://doi.org/10.1177/1049732305276687>

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology, 3*, 29-51. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091419>

* Kaufman, A. S., & Lichtenberger, E. O. (2005). *Assessing adolescent and adult intelligence*. John Wiley & Sons.

* Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005). *Essentials of KABC-II assessment*. Hoboken, Wiley.

* Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V*. Wiley.

Kranzler, J. H., Benson, N., & Floyd, R. G. (2016). Intellectual assessment of children and youth in the United States of America: Past, present, and future. *International Journal of School & Educational Psychology, 4*(4), 276-282. <https://doi.org/10.1080/21683603.2016.1166759>

* Kranzler & Floyd (2020). *Assessing intelligence in children and adolescents: A practical guide for evidenced-based assessment*. Rowman and Littlefield.

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 learning disabilities?. *Contemporary School Psychology*. <https://doi.org/10.1007/s40688-020-00274-0>

Kuhn, T. S. (1967). *The structure of scientific revolutions*. University of Chicago Press.

* Lichtenberger E. O. & Kaufman, A. S. (2007) The assessment of preschool children with the Kaufman Assessment Battery for Children, Second Edition (KABC-II). In B. A. Bracken & R. J. Nagle (Eds.), *Psychoeducational Assessment of Preschool Children* (4th ed.). Erlbaum Associates Publishers.

* Lichtenberger, E. O., Sotelo-Dynega, M., & Kaufman, A. S. (2009). *The Kaufman Assessment Battery for Children—Second Edition*. In J. A. Naglieri & S. Goldstein, *Practitioner's guide to assessing intelligence and achievement* (p. 61–93). John Wiley & Sons Inc.

- Lilienfeld, S. O., Ammirati, R. J., & 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.
- Lilienfeld, S. O., Lynn, S. J., O'Donohue, W. T., & Latzman, R. D. (2017). Epistemic humility: An overarching educational philosophy for clinical psychology programs. *The Clinical Psychologist, 70*(2), 6-14.
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2006). Why questionable psychological tests remain popular. *Scientific Review of Alternative Medicine, 10*, 6-15.
- Lockwood, A. B., & Farmer, R. L. (2019). The cognitive assessment course: Two decades later. *Psychology in the Schools, 57*(2), 265-283. <https://doi.org/10.1002/pits.22298>
- McGill, R. J. (2017). Invalidating the full scale IQ score in the presence of significant factor score variability: Clinical acumen or clinical illusion?. *Archives of Assessment Psychology, 6*(1), 49-79.
- 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>
- McGill, R. J., Ward, T. J., & Canivez, G. L. (2020). Use of translated and adapted versions of the WISC-V: Caveat emptor. *School Psychology International, 41*(3), 276-294. <https://doi.org/10.1177/0143034320903790>
- McGrew, K. S. (2018, April 12). *Dr. Kevin McGrew and Updates to CHC Theory* [Video webcast]. Invited presentation for School Psyched! <https://itunes.apple.com/us/podcast/episode-64-dr-kevin-mcgrew-and-updates-to-chc-theory/id1090744241?i=1000408728620&mt=2>
- * McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). Technical Manual. Woodcock-Johnson IV. Itasca, IL: Riverside.

McKay, V. R., Morshed, A. B., Brownson, R. C., Proctor, E. K., & Prusaczyk, B. (2018). Letting go:

Conceptualizing intervention de-implementation in public health and social service settings.

American Journal of Community Psychology, 62(1-2), 189–202.

<https://doi.org/10.1002/ajcp.12258>

McNamara, K. M., Walcott, C. M., & Hyson, D. (2019). *Results From the NASP 2015 Membership Survey*,

Part Two: Professional Practices in School Psychology [Research report]. Bethesda, MD: National

Association of School Psychologists.

* Miller, D. C. & Maricle, D. E. (2019). *Essentials of school neuropsychological assessment* (3rd ed.). Wiley

& Sons.

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>

Meichenbaum, D., & Lilienfeld, S. O. (2018). How to spot hype in the field of psychotherapy: A 19-item

checklist. *Professional Psychology: Research and Practice*, 49(1), 22–30.

<https://doi.org/10.1037/pro0000172>

Miller, L. T., Bumpus, E. C. & Graves, S. L. (2020). The state of cognitive assessment training in school

psychology: An analysis of syllabi. *Contemporary School Psychology*.

<https://doi.org/10.1007/s40688-020-00305-w>

Oakland, T. D., & Zimmerman, S. A. (1986). The course on individual mental assessment: A national

survey of course instructors. *Professional School Psychology*, 1(1), 51.

* Roid, G. H., & Barram, R. A. (2004). *Essentials of Stanford-Binet intelligence scales (SB5) assessment*

(Vol. 39). John Wiley & Sons.

* Sattler, J. M. (2018). *Assessment of children: Cognitive foundations and applications* (6th ed.). La

Mesa, CA: Jerome M. Sattler Publisher.

- 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* (p. 73–163). The Guilford Press.
- Schneider, W. J., & Roman, Z. (2017). Fine-tuning cross-battery assessment procedures: After follow-up testing, use all valid scores, cohesive or not. *Journal of Psychoeducational Assessment, 36*(1), 34-54. <https://doi.org/10.1177/0734282917722861>
- * Schrank, F. A., Decker, S. L., & Garruto, J. M. (2016). *Essentials of WJ IV cognitive abilities assessment*. John Wiley & Sons.
- * Schrank, F. A., & Wendling, B. J. (2018). *The Woodcock–Johnson IV: Tests of cognitive abilities, tests of oral language, tests of achievement*. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (p. 383–451). The Guilford Press.
- Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools, 51*(10). <https://doi.org/10.1002/pits.21802>
- Truscott, S. D., Baumgart, M. B., & Rogers, K. M. (2004). Financial conflicts of interest in the school psychology assessment literature. *School Psychology Quarterly, 19*(2), 166-178.
- Warne, R. T., Astle, M. C., & Hill, J. C. (2018). What do undergraduates learn about human intelligence? An analysis of introductory psychology textbooks. *Archives of Scientific Psychology, 6*(1), 32-50. <http://dx.doi.org/10.1037/arc0000038>
- Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly, 15*(4), 465–479. <https://doi.org/10.1037/h0088802>
- Watkins, M. W. (2003). IQ subtest analysis: Clinical acumen or clinical illusion?. *The Scientific Review of Mental Health Practice: Objective Investigations of Controversial and Unorthodox Claims in Clinical Psychology, Psychiatry, and Social Work, 2*(2), 118-141.

* Wechsler, D. (2014). Wechsler intelligence scale for children (Fifth Edition.). San Antonio, TX: NCS
Pearson

Youngstrom, E. A. (2013). Future directions in psychological assessment: Combining evidence-based
medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical
Child and Adolescent Psychology*, 42(1), 139–159.

<https://doi.org/10.1080/15374416.2012.736358>

Table 1*Description of the Classification and Coding Framework Employed in the Present Study*

Code	Description
A	Users are encouraged to focus most, if not all, of their interpretive weight at this level of the test
B	Users are encouraged to consider scores at this level as part of a broader step-by-step <i>levels of analysis</i> approach but their interpretive value is considered superior to scores at other levels of the instrument
C	Users are encouraged to consider scores at this level of the test as part of a broader step-by-step <i>levels of analysis</i> approach but their interpretive value is considered subordinate to scores at other levels of the instrument
D	Scores at this level should be interpreted with caution and/or used only for generating clinical hypotheses to be corroborated by other sources of data
E	Interpretation at this level of the test is not encouraged
F	The validity or <i>meaningfulness</i> of a score may be called into question in the presence of significant scatter
N/A	Specific interpretive guidance relative to this level of the test could not be located
Stratum III	FSIQ, global composite, or equivalent
Stratum II	Broad ability index and composite scores
Stratum I	Subtests or measures of narrow abilities
Ancillary ^a	Additional Stratum II level scores that are not derived from factor analysis (i.e., pseudo composites)
Item (Yes, No, N/A)	Users are encouraged to evaluate an examinee's performance on individual items
Behavior (Yes, No, N/A)	Users are encouraged to generate inferences based on their observation of test session behaviors

Note. The levels of analysis approach (e.g., Kaufman, 1994) generally encourages the clinicians to interpret scores in a step-wise fashion beginning with Stratum I and culminating at Stratum III.

^a If available

Table 2

Interpretive Guidance for the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V; Wechsler, 2014)

Resource	Stratum III	Stratum II	Stratum I	Ancillary	Item	Behavior
Flanagan & Alfonso (2017)	C ^a	B ^a	E	C ^a	No	Yes
Groth-Marnat & Wright (2016)	C, F	B, F	C	C, F	Yes	Yes
Kaufman, Raiford, & Coalson (2016)	C	B	C	C	No	Yes
Sattler (2018)	B, F	C, F	C	C, F	Yes	Yes
Wahlstrom et al. (2018)	B	C	C	C	N/A	N/A
Wechsler (2014)*	B	C	C	C	Yes	Yes
Weiss et al. (2016)	B, F	C	C	N/A	No	Yes

Note. Asterisk (*) indicates test technical manual.

^a Although it is suggested that scatter does not impact the validity of a score as a matter of course, users should evaluate the “cohesiveness” of an indicator to determine if the score should be regarded as *clinically meaningful*.

Table 3

Interpretive Guidance for the Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014)

Resource	Stratum III	Stratum II	Stratum I	Ancillary	Item	Behavior
Flanagan & Alfonso (2016)	†	†	†	†	N/A	Yes
McGrew, La Forte, & Schrank (2014)*	†	†	†	†	N/A	Yes
Sattler (2018)	B	C	C	C	Yes	Yes
Schrank, Decker, & Garruto (2016)	C	C	B	C	N/A	Yes
Schrank & Wendling (2018)	†	†	†	†	N/A	N/A

Note. Asterisk (*) indicates test technical manual.

† No specific interpretive guidance is given although it is suggested that interpretive focus will vary depending on the purposes of an evaluation and that all clusters, scores, and tests are interpretable.

Table 4

Interpretive Guidance for the Kaufman Assessment Battery for Children-Second Edition (KABC-II; Kaufman & Kaufman, 2004)

Resource	Stratum III	Stratum II	Stratum I	Ancillary	Item	Behavior
Drozdick et al. (2018)	C	B	N/A	N/A	N/A	Yes
Kaufman et al. (2005)	C, F	B, F	C	C	N/A	Yes
Kaufman & Kaufman (2004)*	C, F	B, F	C	C	N/A	Yes
Lichtenberger & Kaufman (2007)	C, F	B, F	E	C	Yes	Yes
Lichtenberger, Sotelo-Dynega, & Kaufman (2009)	C, F	B, F	E	C	Yes	Yes

Note. Asterisk (*) indicates test technical manual.

Table 5*Interpretive Guidance for the Differential Ability Scales-Second Edition (DAS-II; Elliott, 2007)*

Resource	Stratum III	Stratum II	Stratum I	Ancillary	Item	Behavior
Dumont, Willis, & Elliott (2009)	C, F	B, F	D	C, F	Yes	Yes
Elliott (2007)*	C, F	B, F	C	C, F	N/A	N/A
Elliott et al. (2018)	C, F	B, F	C	C, F	N/A	N/A
Sattler (2018)	C	B	C	C	Yes	Yes

Note. Asterisk (*) indicates test technical manual.

Table 6*Interpretive Guidance for the Stanford-Binet Intelligence Scales-Fifth Edition (SB5; Roid, 2003)*

Resource	Stratum III	Stratum II	Stratum I	Ancillary	Item	Behavior
Alfonso & Flanagan (2007)	B, F	C, F	N/A	C, F	N/A	N/A
Sattler (2018)	A, F	D, F	C	C, F	Yes	Yes
Roid (2003)*	B, F	C, F	C	C, F	Yes	Yes
Roid & Barram (2004)	C, F	C, F	C	B, F	Yes	Yes
Roid & Poplun (2012)	B, F	C, F	C	C, F	Yes	Yes
Roid & Tippin (2009)	B, F	C, F	C	C, F	Yes	Yes

Note. Asterisk (*) indicates test technical manual.

Table 7*Guidance Offered in General Cognitive Assessment Interpretive Texts and Resources*

Resource	Stratum III	Stratum II	Stratum I	Ancillary	Item	Behavior
Canivez (2013)	A	E	E	N/A	N/A	N/A
Flanagan, Ortiz, & Alfonso (2013)	E	A, F	D	D, F	N/A	Yes
Flanagan, Ortiz, Alfonso, & Dynda (2008)	E	A	N/A	C	N/A	Yes
Glutting, Watkins, & Youngstrom (2003)	A	E	E	N/A	N/A	N/A
Hale & Fiorello (2004)	E	A, F	C	N/A	Yes	Yes
Kranzler & Floyd (2013)	A	D	E	N/A	No	Yes
Miller (2013)	N/A	B, F	C	B, F	Yes	Yes

Note. Asterisk (*) indicates test technical manual.