Contemporary School Psychology (2019) 23:194–200
https://doi.org/10.1007/s40688-019-00238-z

Evidence-Based Assessment for School Psychology: Research, Training, and Clinical Practice

Gary L. Canivez [1]

Published online: 12 March 2019

Abstract
Evidence-based assessment (EBA) and evidence-based practice are gaining attention and application in clinical psychology (Hunsley and Mash, Annual Review of Clinical Psychology 3:29–51, 2007; Youngstrom and VanMeter, Clinical Psychology: Science and Practice 23:327–347, 2016) and recently in school psychology. This commentary explores EBA in school psychology with implications for research, training, and clinical practice. Fundamental measurement principles and psychometric knowledge and application are essential for school psychologists to ethically practice and are core elements for EBA. It is hoped that there will be a dramatic increase in diagnostic and treatment utility studies reported in the peer-reviewed literature to help guide EBA in school psychology practice and that school psychology will be substantively guided by EBA despite challenges of inconsistently or inadequately defined disabilities.

Keywords Evidence-based assessment · Evidence-based practice · Measurement · Psychometrics · Ethics

Evidence-based assessment (EBA) and evidence-based practice in psychology (EBPP; American Psychological Association (APA) 2006) are gaining attention and application in clinical psychology (Hunsley and Mash 2007; Youngstrom and Van Meter 2016) and recently, more formally introduced and encouraged in school psychology. Greater application has been seen in medicine where it originated to optimize decision-making using best available and updated empirical evidence to improve patient outcomes (Hunsley and Mash 2007; Sackett et al. 1996; Youngstrom 2013). Greater use in medicine may be related to the fact that diagnoses in medicine may have biological markers to identify presence of conditions, which many psychological disorders lack. Still, EBA has much to offer in application to psychology broadly (APA 2006) and school psychology in particular. Many of the approaches suggested offer improved practices but there are also challenges (Reynolds 2016). While EBA may seem to be a new approach, numerous underlying tenets have actually been around since the development, evaluation, and use of the first psychological tests. There have also been applications of sophisticated EBA methods related to actuarial decision-making approaches combining various measures such as with the McDermott Multidimensional Assessment of Children (M-MAC) (McDermott and Watkins 1985) as well as methods for interpreting individuals' similarity to normative profiles within a specific test to aid in differential diagnosis as performed with the Adjustment Scales for Children and Adolescents (McDermott 1993, 1994; McDermott et al. 1993; McDermott and Weiss 1995). Widespread systematic use of EBA, however, does not seem to be the norm and as a result, much of school psychology practice may at present best be characterized by application of clinical judgment.

As is frequently discussed in introductory texts in psychological measurement and assessment, testing is a more specific term related to the administration of a measure or procedure to quantify a psychological construct, while assessment is broader and focused on multiple methods in order to answer specific referral questions. Assessment in school psychology is broadly conceived and includes numerous methods and data sources (observation, interview, case study, and psychological/educational tests). Within the assessment domain, the use of tests and procedures to aid in psychodiagnosis or classification and treatment recommendations are guided by specific principles prescribed by the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on

✉ Gary L. Canivez
glcanivez@eiu.edu

[1] Department of Psychology, Eastern Illinois University, 600 Lincoln Avenue, Charleston, IL 61920-3099, USA

Measurement in Education (AERA, APA, NCME) 2014) to facilitate understanding what evidence is needed to support various interpretations and test uses. NASP training standards (NASP 2010a) and standards of practice (NASP 2010b) place great emphasis on Data-Based Decision-Making, which links good decisions to quality data and is particularly important in high-stakes decisions. Further, there are general and specific ethical principles (APA 2002, 2010; NASP 2010c) that address test score use and interpretations that guide assessment practice. Weiner's (1989) sage advice is that effective psychodiagnosticians:

> (a) know what their tests can do and (b) act accordingly. Knowing what one's test can do—that is, what psychological functions they describe accurately, what diagnostic conclusions can be inferred from them with what degree of certainty, and what kinds of behavior they can be expected to predict—is the measure of a psychodiagnostician's competence. Acting accordingly—that is, expressing only opinions that are consonant with the current status of validity data—is the measure of his or her ethicality (p. 829).

EBA is implicitly guided by these principles because applications of information derived from tests or procedures must be empirically supported to be included and only information that has been shown to substantively reduce uncertainty in clinical decisions is considered useful.

Definitions of Evidence-Based Assessment

Hunsley and Mash (2007) defined EBA as "an approach to clinical evaluation that uses research and theory to guide the selection of constructs to be assessed for a specific assessment purpose, the methods and measures to be used in the assessment, and the manner in which the assessment process unfolds" (p. 30). They indicated that EBA was a necessary precursor to evidence-based *treatment* within EBPP, which APA (2006) defined as "the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences" in order "to promote effective psychological practice and enhance public health by applying empirically supported principles of psychological assessment, case formulation, therapeutic relationship, and intervention" (p. 273). What is specified in these definitions is that information provided by assessment procedures or tests must be empirically supported and produce improved client outcomes.

History/Roots of Evidence-Based Assessment

EBA has strong roots in seminal works of Paul Meehl (Meehl 1954; Meehl and Rosen 1955), Robyn Dawes (Dawes, 2005; Dawes et al. 1989), Richard McFall (1991, 2000, 2005), and their colleagues (Faust 1986, 1990; Garb 2005; Grove and Meehl 1996; Grove et al. 2000; Lilienfeld, Wood, & Garb, 2006). School psychologists frequently use test scores, score comparisons, and analysis information as well as test session observations, background information, interview information, etc. to make clinical decisions, judgments, or inferences regarding students. Such decisions might be considered *clinical* decisions rather than *actuarial* (statistical) decisions because it is the clinician's *judgment* of meaning of the score(s) that guides their interpretation (decision) rather than strict adherence to a statistically based (formula) interpretation (decision) (Meehl 1954, 1957; Meehl and Rosen 1955).

Traditional test interpretation practices and clinical decision-making rely on the clinician's judgment regarding assessment data gathered and there are numerous errors in judgment that could result including confirmation bias, overconfidence, fundamental attribution error, misperception of regression, representativeness, insensitivity to prior probabilities (base rates), misperception about chance (i.e., illusory correlations, conjunction fallacy, inverse probabilities, insensitivity to sample size (law of small numbers), pseudodiagnosticity), and hindsight bias (Garb 1997, 1998, 2005; Kahneman et al. 1982; Lilienfeld et al. 2012; Meehl and Rosen 1955; Tversky & Kahneman, 1974; Watkins 2009). Inconsistent application of diagnostic criteria (decision rules), inconsistent weighting of diagnostic variables, and inconsistent decision-making processes (strategies or sequences) were noted problems among school psychologists identified by McDermott (1981). EBA methods help to minimize these potential errors.

As noted by Canivez (2013), over 60 years ago, Paul Meehl launched a debate on actuarial decision-making by inquiring about the relationship between clinical prediction and actuarial (statistical) prediction (Meehl 1954), which he referred to as his "wicked book" (Meehl 1979, p. 564) or "disturbing little book" (Meehl 1986, p. 370). Meehl concluded that the actuarial approach was superior and as such should be more frequently used. Numerous studies comparing clinical (informal or impressionistic) and actuarial (formal, mechanical, algorithmic) predictive methods since Meehl's seminal work has fairly consistently shown that the actuarial method was generally observed to be as accurate or more accurate than clinical methods (Dawes et al. 1989; Grove and Meehl 1996; Grove et al. 2000). While most studies in the Grove et al. meta-analysis identified statistical equivalence between the clinical and actuarial methods, it is argued that there should be preference for the actuarial method in the event of a tie, because once developed it is more cost and time effective, less laborious, and allows for consistent, dispassionate application (Dawes et al. 1989; Meehl 1954). Statistical methods used in actuarial approaches (multiple regression, logistic regression, discriminant function analysis) differentially and optimally weight predictor variables to provide the

most accurate prediction of the criterion variable which is a distinct advantage over that of a clinician (Grove and Meehl 1996). It is for these and other reasons that Grove and Meehl argued that actuarial methods should be widely applied and false arguments against actuarial methods be rejected.

Forms of Evidence-Based Assessment

At the foundational level, all EBA depend on measures or procedures that produce scores or information that have strong evidence for reliability, validity, and utility, and these are hierarchically related with reliability, a foundation without which validity is not possible. Validity relates to inferences made from scores and score comparisons that are necessary, but not sufficient, for utility. Utility generally relates to the observation that clients have improved in their functioning as a result of services provided based on the assessment data (Hunsley and Mash 2007). There are several aspects within utility that are sometimes examined separately and include *diagnostic utility* (the correct identification of those who truly have a condition and those who truly do not have a condition) and *treatment utility* (that the assessment information resulted in recommendation of a specific treatment that as a result improved the client's functioning). From a diagnostic utility perspective, EBA would involve identifying and using variables or comparisons that aid in ruling in some condition (positive predictive power) or ruling out some condition (negative predictive power).

Thus, psychometric studies of tests and procedures assessing reliability and validity of measurement provide evidence for the *potential* for clinical utility. As such, basic reliability and validity studies will help determine which scores or procedures are inadequate and thus unlikely to provide utility (and ought not to be used in clinical decision-making). For scores or procedures, which show adequate reliability and validity, they can then be examined for utility. Studies using discriminant function analysis or logistic regression with follow-up assessment of diagnostic efficiency (positive predictive power, negative predictive power, sensitivity, specificity, false positive, and false negative rates (Kessel and Zimmerman 1993)) and receiver operating characteristic (ROC) curve analyses (McFall and Treat 1999; Treat and Viken 2012) are useful in determining how well the test scores or variables perform in correctly classifying individuals. The benefit of ROC analyses is that the area under the curve (AUC) is not affected by base rates or cut scores (McFall 2005). Information from these analyses provides estimates of likelihood ratios useful in determining the posterior probability of diagnostic classification or outcome. Variables (test scores or combinations of test scores) included will change the "risk" or "diagnosis" previously estimated by the antecedent (prior) probability (usually a base rate). The use of nomograms, spreadsheet calculators, or web-based calculators allow users to enter prior probabilities and likelihood ratios from assessment information to produce posterior probabilities without needing to do the Bayesian math of combining probabilities satisfying Bayes Theorem (Bayes 1763). Several articles by Youngstrom and colleague (Youngstrom 2013, 2014; Youngstrom and Van Meter 2016) describe in detail such applications that have previously been applied in EBA in clinical psychology. Cluster analysis (CA), conceptually the opposite of factor analysis as it is involved in identifying groups of similar *people*; latent class analysis (LCA), identifying population subgroups with categorical variables; or latent profile analysis (LPA), identifying population subgroups with continuous variables are helpful in determining the number of distinct groups and their characteristics on important variables (Flaherty and Kiff 2012; McDermott 1998; McDermott and Weiss 1995). Knowledge of latent groups or clusters allows for the determination of similarity of an individual to the latent group and empirically based classification or diagnosis.

Importance of Evidence-Based Assessment

EBA is important as noted above because the goals are to provide better and empirically supported assessments that would provide better outcomes for students. Ultimately, EBA should produce better diagnostic or classification utility (accuracy) *and* lead to specific treatments that improve student functioning (treatment utility). By focusing on empirically supported variables or scores in the assessment process, school psychologists would be less likely to be victims of the various cognitive errors that negatively affect decision-making and scientific thinking (Lilienfeld et al. 2012; Watkins 2009). Another benefit of EBA is that by knowing what data are useful for ruling in or ruling out conditions, assessment can be more efficient and targeted. When including unreliable or irrelevant information in the assessment process, predictions or classifications will suffer from a "dilution effect" whereby predictions are less accurate (Nisbett et al. 1981). Thus, test scores and comparisons that are unsatisfactory in reliability or validity estimates ought not to be considered and this has implications for test selection and test score reporting (i.e., reporting only scores and comparisons that are psychometrically sound). Another benefit is that limitations of human cognitive abilities in problem solving or decision-making (i.e., inability to differentially weight variables, limited capacity for thinking about multiple variables, and their complex interactions) are minimized by using EBA. Youngstrom and Van Meter (2016) noted IBM Watson is providing such advances in medicine and similar approaches could be developed for school psychology. McDermott and Watkins (1985) produced a program three decades ago that provided similar EBA that was well ahead of its time and now could be

housed in a smartphone app. Such approaches could also take advantage of new updates in research that may be integrated into the system to quickly make additional improvements.

Research

Research necessary for supporting EBA is prescribed by traditional and modern methods of applied psychometrics in the assessment of reliability, validity, and utility, and these provide the psychometric information that provides empirical support for tests or procedures required for application in EBA. Thus, empirical evidence for support of reliability may be generated from internal consistency, test-retest (stability), alternate forms (equivalence), and/or interrater agreement estimates. Some of these may be more important depending on the purpose for the score and decision. Item response theory metrics may also provide useful information (Embretson and Reise 2000). Empirical evidence for validity was formerly conceptualized according to the Trinitarian model (content, criterion-related, construct), but Messick's unified validity theory (Messick 1995) provides a contemporary approach that focuses on evidence to support specific interpretations for intended purposes. Messick's approach is incorporated into the *Standards for Educational and Psychological Testing* (AERA, APA, NCME 2014) and considers all validity as construct validity, but based on evidence from five sources: (1) test content, (2) response processes, (3) internal structure, (4) relations with other variables, and (5) consequences of testing (social consequences), but this last source is contentious. Assessment of test score validity might be supported by various methods such as distinct group differences, exploratory and confirmatory factor analyses, convergent and discriminant validity, and predictive validity. In the case of distinct group differences, statistically significant group differences would be necessary, but not sufficient, for utility, which should be routinely examined and reported. Utility studies may be supported by examinations of diagnostic utility (accuracy) and/or treatment utility. Empirical evidence from these approaches is required and is ultimately the most valuable metrics to justify use in EBA. EBA also recognizes that empirical results are specific to the particular study, which includes a particular sample from a particular region, with a particular method, and replication and cross-validation is necessary.

Training

Preparation of school psychologists for clinical practice and/or faculty/trainer positions is a lengthy process, some of which is devoted to psychological assessment in order that professionals may competently administer, score, and interpret their tests or procedures. Turner et al. (2001) presented guidelines required for competent test use, including psychometric and measurement knowledge with elaborate lists of subtopics related to reliability (and measurement error), validity (and test score meaning), norms (normative interpretation), and selection of tests for use. However, surveys of graduate training in statistics and measurement by Aiken, West, Sechrest, and Reno (1990) and Aiken, West, and Millsap (2008) noted inadequate training requirements at the doctoral level. Aiken et al. (2008) noted "we find it deplorable that a dozen years later, the measurement requirement occupies a median of only 4.5 weeks in the PhD curriculum in psychology" (p. 43). They further lamented the inadequacies of measurement training that resulted in a situation "that most graduates lacked fundamental competency in measurement" (p. 43). With respect to school psychology programs, there is likely great variability in required coursework and emphasis on advanced statistics and psychological measurement, but given the importance for clinical assessment and high-stakes decisions, measurement training should be prominently featured in *all* programs.

This training problem has its origins in educational curricula for psychology majors at the undergraduate level where, while it is noted that "students will: 2.4E design and adopt high-quality measurement strategies that enhance reliability and validity" (APA 2013, p. 22), this author knows of no US universities that *require* a psychological measurement course but universally require at least one statistics course and one experimental design or research methods course. Psychological measurement (applied statistics and procedures for assessment of reliability, validity, and diagnostic utility) is the third leg of a metaphorical scientific foundations stool and by not requiring such a course, the discipline implicitly tells students measurement may not be so important. At the graduate level, there do not appear to be specific requirements for advanced psychological measurement training that would adequately prepare students and the consequences pointed out by Aiken et al. (2008) are all too common. However, APA (2016) Section C: Implementing Regulations (IRs) of the Standards of Accreditation now lists psychometrics as part of Category 4 of discipline-specific knowledge (Standards of Accreditation (Doctoral Standards, II.B.1.a.)), although the method for achieving this discipline-specific knowledge is left to programs. In the 12 + years, this author has taught continuing education workshops on psychological measurement principles and ethical implications in the USA and abroad, most attendees comment that they received little direct training that would prepare them to adequately evaluate psychometric fitness of the tests they frequently use. While Turner et al. (2001) specified qualifications of test users, without formal advanced courses in psychological measurement and statistics, it is unlikely that school psychologists would be satisfactorily prepared to evaluate test technical qualities reported in manuals or to be able to independently judge the empirical evidence for test score reliability, validity, and utility in the professional peer-reviewed literature.

The lack of adequate training in psychological measurement principles and statistical techniques to assess reliability, validity, and utility means that many school psychologists may well be incapable of identifying what *is not* included in test technical manuals (but should be) or possibly worse be unable to identify what in a test technical manual is inadequately reported or conclusions exaggerated or made without sufficient empirical evidence. Each test interpretation method specified in test manuals are required to have evidence for psychometric fitness in order to be used as prescribed (AERA, APA, NCME 2014) and a foundation for EBA. However, many technical manuals still do not provide adequate evidence of empirical support for all their recommended interpretations, and in some cases, recommended interpretation practices are counter to what volumes of empirical research in the peer-reviewed literature show. This is clearly documented with the continued recommendation of ipsative subtest comparisons and other cognitive profile analysis methods in intelligence tests (McGill, Dombrowski, & Canivez, 2018). Oscar Buros noted in the *Preface* of the first *Mental Measurements Yearbook* (MMY):

> Test users have every right to demand that test authors and publishers present full particulars concerning the methods used in constructing and validating the tests which they place on the market (Buros 1938, p. 13).

Presently, there are numerous examples of test technical manuals that 80 years later, still fall far short of this call. Without adequate measurement training, how would a school psychologist know? EBA requires evidence of psychometric fitness of tests or procedures used. In order for school psychologists to be capable of EBA, they need to be properly educated in advanced psychological measurement and statistics to adequately assess test technical manuals, MMY, and the extant literature, including recognizing information generated by those with financial conflict of interest who may be promoting tests or procedures that are not sufficiently empirically supported.

Clinical Practice/Implementation

It was pointed out by Youngstrom (2013) that surveys of clinical psychology assessment practices of clinicians illustrate that they use the tests and procedures that they were specifically taught and often do not deviate from later. Thus, if EBA is to become a standard practice, graduate students will need to be specifically taught EBA approaches. A related aspect is that many training programs promote a *scientist-practitioner* model. If school psychologists are to be scientist practitioners, they should use and report scores and comparisons that are sufficiently empirically supported, and they ought to modify or abandon assessment and interpretation practices when empirical research demonstrates the unreliability or dubious validity of scores, score comparisons, or procedures that generate information to be used in assessment. EBA specifically updates methods, tests, scores, etc. with the most recent empirical evidence, which also helps to keep assessment current and relevant.

As noted above, there are many advantages that EBA provides with the ultimate benefit of better practices and services to students. EBA methods would also include the benefit of consistency of application. In school psychology application, there are added challenges due to assessments that interface with individual states laws and regulations for implementation of IDEA. In states, where local school districts may be allowed to use various and differing procedures, it may also be as problematic to obtain consistency between school districts as it is between states.

Looking Forward

This special issue presents several studies examining psychometric fitness of various tests or score comparisons influencing utility. Given the role that assessment continues to play in school psychology, it is hoped that such studies will increase exponentially and address the critical questions of what scores, comparisons, or procedures produce viable information to accurately rule in or rule out conditions and provide necessary direction for effective instruction or treatment.

References

Aiken, L. S., West, S. G., Sechrest, L. B., & Reno, R. R. (1990). Graduate training in statistics, methodology, and measurement in psychology: an international survey. *American Psychologist, 45*, 721–734. https://doi.org/10.1037/0003-066X.45.6.721.

Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist, 63*, 32–50. https://doi.org/10.1037/0003-066X.63.1.32.

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

American Psychological Association. (2002, 2010 Amendments). *Ethical principles of psychologists and code of conduct.* Washington, DC: APA.

American Psychological Association. (2013). *APA guidelines for the undergraduate psychology major: Version 2.0*. Washington, DC: APA.

American Psychological Association. (2016). Implementing Regulations Section C: IRs related to the Standards of Accreditation. Retrieved from https://www.apa.org/ed/accreditation/section-c-soa.pdf.

American Psychological Association, Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61*, 271–285. https://doi.org/10.1037/0003-066X.61.4.271.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical transactions of the Royal Society, 53*, 370–418 Reprinted in G. A. Barnard (1958), Studies in the history of probability and statistics. Biometrika, 45, 293–315.

Buros, O. K. (1938). Preface. In O. K. Buros (Ed.), *The 1938 mental measurements yearbook* (p. xiii). Oxford: Rutgers University Press.

Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessments* (pp. 84–112). New York: Oxford University Press.

Dawes, R. M. (2005). The ethical implications of Paul Meehl's work on comparing clinical versus actuarial prediction methods. *Journal of Clinical Psychology, 61*, 1245–1255. https://doi.org/10.1002/jclp.20180.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668–1674. https://doi.org/10.1126/science.2648573.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice, 17*, 420–430. https://doi.org/10.1037/0735-7028.17.5.420.

Faust, D. (1990). Data integration in legal evaluations: Can clinicians deliver on their premises? Behavioral Sciences and the Law, 7, 469–483. https://doi.org/10.1002/bsl.2370070405.

Flaherty, B. P., & Kiff, C. J. (2012). Latent class and latent profile models. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *Handbook of research methods in psychology: volume 1 foundations, planning, measures, and psychometrics* (pp. 391–404). Washington, DC: American Psychological Association. https://doi.org/10.1037/13621-019.

Garb, H. N. (1997). Race bias, social class bias, and gender bias in clinical judgment. *Clinical Psychology: Science and Practice, 4*, 99–120. https://doi.org/10.1111/j.1468-2850.1997.tb00104.x.

Garb, H. N. (1998). *Studying the clinician: judgment research and psychological assessment*. Washington, DC: American Psychological Association. https://doi.org/10.1037/10299-000.

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67–89. https://doi.org/10.1146/annurev.clinpsy.1.102803.143810.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical–statistical controversy. *Psychology, Public Policy, and Law, 2*, 293–323. https://doi.org/10.1037/1076-8971.2.2.293.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta–analysis. *Psychological Assessment, 12*, 19–30. https://doi.org/10.1037/1040-3590.12.1.19.

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.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511809477.

Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self-administered questionnaires: extent of the problem, recommendations for standardized presentation of results, and implications for the peer review process. *Psychological Assessment, 5*, 395–399. https://doi.org/10.1037/1040-3590.5.4.395.

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*, 7–36. https://doi.org/10.1016/j.jsp.2011.09.006.

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2006). Why questionable psychological tests remain popular. The Scientific Review of Alternative Medicine, 10, 6–15.

McDermott, P. A. (1981). Sources of error in psychoeducational diagnosis of children. *Journal of School Psychology, 19*, 31–44. https://doi.org/10.1016/0022-4405(81)90005-4.

McDermott, P. A. (1993). National standardization of uniform multisituational measures of child and adolescent behavior pathology. *Psychological Assessment, 5*, 413–424. https://doi.org/10.1037/1040-3590.5.4.413.

McDermott, P. A. (1994). *National profiles in youth psychopathology: manual of adjustment scales for children and adolescents*. Philadelphia, PA: Edumetric and Clinical Science.

McDermott, P. A. (1998). MEG: Megacluster analytic strategy for multistage hierarchical grouping with relocations and replications. *Educational and Psychological Measurement, 58*, 677–686. https://doi.org/10.1177/0013164498058004010.

McDermott, P. A., & Watkins, M. W. (1985). *Microcomputer systems manual for McDermott multidimensional assessment of children*. New York: The Psychological Corporation.

McDermott, P. A., & Weiss, R. V. (1995). A normative typology of healthy, subclinical, and clinical behavior styles among American children and adolescents. *Psychological Assessment, 7*, 162–170. https://doi.org/10.1037/1040-3590.7.2.162.

McDermott, P. A., Marston, N. C., & Stott, D. H. (1993). *Adjustment scales for children and adolescents*. Philadelphia, PA: Edumetric and Clinical Science.

McFall, R. M. (1991). Manifesto for a science of clinical psychology. *The Clinical Psychologist, 44*, 75–88.

McFall, R. M. (2000). Elaborate reflections on a simple manifesto. *Applied and Preventive Psychology, 9*, 5–21. https://doi.org/10.1016/S0962-1849(05)80035-6.

McFall, R. M. (2005). Theory and utility—key themes in evidence–based assessment: comment on the special section. *Psychological Assessment, 17*, 312–323. https://doi.org/10.1037/1040-3590.17.3.312.

McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessment with signal detection theory. *Annual Review of Psychology, 50*, 215–241. https://doi.org/10.1146/annurev.psych.50.1.215.

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.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and review of the evidence*. Minneapolis, MN: University of Minnesota Press. https://doi.org/10.1037/11281-000.

Meehl, P. E. (1957). When shall we use our heads instead of the formula? Journal of Counseling Psychology, 4, 268–273. https://doi.org/10.1037/h0047554.

Meehl, P. E. (1979). A funny thing happened to us on the way to latent entities. *Journal of Personality Assessment, 43*, 564–581. https://doi.org/10.1207/s15327752jpa4306_2.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50*, 370–375. https://doi.org/10.1207/s15327752jpa5003_6.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194–216. https://doi.org/10.1037/h0048070.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749. https://doi.org/10.1037/0003-066X.50.9.741.

National Association of School Psychologists. (2010a). *Standards for graduate preparation of school psychologists*. Bethesda, MD: NASP.

National Association of School Psychologists. (2010b). *Model for comprehensive and integrated school psychological services*. Bethesda, MD: NASP.

National Association of School Psychologists. (2010c). *Principles of professional ethics*. Bethesda, MD: NASP.

Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology, 12*, 248–277. https://doi.org/10.1016/0010-0285(81)90010-4.

Reynolds, C. R. (2016). Contextualized evidence and empirically based testing and assessment. *Clinical Psychology: Science and Practice, 23*, 410–416. https://doi.org/10.1111/cpsp.12181.

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal, 312*, 71–72. https://doi.org/10.1136/bmj.312.7023.71

Treat, T. A., & Viken, R. J. (2012). Measureing test performance with signal detection theory techniques. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *Handbook of research methods in psychology: volume 1 foundations, planning, measures, and psychometrics* (pp. 723–744). Washington, DC: American Psychological Association.

Turner, S. M., DeMers, S. T., Fox, H. R., & Reed, G. M. (2001). APA's guidelines for test user qualifications: an executive summary. *American Psychologist, 56*, 1099–1113. https://doi.org/10.1037/0003-066X.56.12.1099.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124–1131. https://doi.org/10.1126/science.185.4157.1124.

Watkins, M. W. (2009). Errors in diagnostic decision making and clinical judgment. In T. B. Gutkin & C. R. Reynolds (Eds.), *Handbook of school psychology* (4th ed., pp. 210–229). Hoboken, NJ: Wiley.

Weiner, I. B. (1989). On competence and ethicality in psychodiagnostic assessment. *Journal of Personality Assessment, 53*, 827–831. https://doi.org/10.1207/s15327752jpa5304_18.

Youngstrom, E. A. (2013). Future directions in psychological assessment: combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. I139–159. https://doi.org/10.1080/15374416.2012.736358.

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *Journal of Pediatric Psychology, 39*, 204–221. https://doi.org/10.1093/jpepsy/jst062.

Youngstrom, E. A., & Van Meter, A. (2016). Empirically supported assessment of children and adolescents. *Clinical Psychology: Science and Practice, 23*, 327–347. https://doi.org/10.1111/cpsp.12172.

Gary L. Canivez Ph.D. is Professor of Psychology at Eastern Illinois University and principally involved in the Specialist in School Psychology program. Before entering academia Dr. Canivez was a school psychologist for 8 years, was on the adjunct faculty of Arizona State University and Northern Arizona University, and was president of the Arizona Association of School Psychologists. Dr. Canivez currently serves as Associate Editor of Archives of Scientific Psychology and served as Associate Editor for Psychological Assessment. He is a consulting editor and frequent reviewer for numerous other school psychology, assessment, and clinically oriented journals. Dr. Canivez is a member of the Society for the Study of School Psychology, a Fellow of the Division (5) of Quantitative and Qualitative Methods and Division (16) of School Psychology of the American Psychological Association, and a Charter Fellow of the Midwestern Psychological Association. The author of over 100 research and professional publications and over 200 professional presentations and continuing professional development workshops, Dr. Canivez has research interests in psychological assessment and measurement pertaining to intelligence, achievement, personality, and psychopathology; and his research has been supported by the National Institutes of Health/National Institute of Mental Health.