Warning Signs for Hype in School-Based Assessment:
Implications for Training and Pedagogy

Ryan L. Farmer, Ph.D.
Oklahoma State Univeristy

Ryan J. McGill, Ph.D.
William & Mary

Adam B. Lockwood, Ph.D.
Kent State University

Stefan C. Dombrowski, Ph.D.
Rider University

Gary L. Canivez, Ph.D.
Eastern Illinois University

Imad Zaheer, Ph.D.
St Johns University

Abstract: This article addresses the use of hype in the promotion of clinical assessment practices and instrumentation. Particular focus is given to the role of school psychologists in evaluating the evidence associated with clinical assessment claims, the types of evidence necessary to support such claims, and the need to maintain a degree of "healthy self-doubt" about one's own beliefs and preferred practices. Included is a discussion of topics that may facilitate developing and refining scientific thinking skills related to clinical assessment across common coursework, and how this framework fits with both the scientist-practitioner and clinical science perspectives for training..

Keywords: assessment; hype; scientific thinking; epistemology; evidence

The importance of evidence-based assessment (American Psychological Association [APA], 2006, 2017) in school psychology is well recognized. This framework assists with accurate decision-making related to diagnosis and treatment selection in a variety of practice settings (e.g., clinics, schools). In addition, using scientifically-grounded practices may mitigate the risk for various types of harm (Lilienfeld, 2016; Lilienfeld et al., 2019), including, but not limited to, allocating scarce resources and time, and inefficiently and inadvertently relying on inadequate information that may lead to misdiagnosis. The latter, in particular, may lead to unnecessary denial of needed services, placement in an overly restrictive therapeutic environment, and/or the provision of practices that may be ineffective or, in some cases, cause harm. McFall (1991, 1996, 2000) advocated that integrating science into psychological practice meant that claims associated with any given practice should be thoroughly scrutinized including weighing the risk of the potential negative side-effects against anticipated positive gains—what Lilienfeld and colleagues (2019) considered the underlying rationale of the evidence-based practice (EBP) movement. Clinical practices that fall outside of what evidence supports have been referred to as low-value practices (LVPs) within the professional literature, because they have not been shown to be beneficial, are less effective than other available alternatives, or are therapeutically unnecessary and/or contraindicated by research evidence (McKay et al., 2018). In contrast to LVPs, Pratkanis (1995) and others (e.g.,

Meichenbaum & Lilienfeld, 2018; Travers, 2016) refer to pseudoscience or pseudoscientific practices as another problematic class deserving consideration in these discussions. Generally, whereas demarcation between science, non-science, and pseudoscience is made upon the dual lines of theoretical understanding and empirical evidence (Pigliucci, 2018), LVPs are primarily focused on the functional outcomes of the practice itself (see McKay et al., 2018). This sets LVP in direct contrast with EBP, and pseudoscientific practices in direct contrast with science-based practice (Lilienfeld et al., 2018). Pseudoscience is best understood in contrast with science (Pigliucci, 2018). For simplicity, we only refer to LVP in this manuscript to emphasize the importance of clinical outcomes. Despite the risks associated with using untested and questionable approaches, assessment practices in school psychology are muddled with LVPs (e.g., Allen & Hanchon, 2013; Benson et al., 2019, 2020; Gross et al., 2019; Kranzler et al., 2020; Silva et al., 2020; Sotelo-Dynega & Dixon, 2014).

What maintains the use of LVPs is not well understood. There are likely several contributing factors such as the dissemination and promotion of selected LVPs in graduate education (J. Cook et al., 2009), in textbooks and interpretive guidebooks (Farmer et al., 2020), and through continuing education sources (Washburn et al., 2019). Other factors such as outdated, inconsistent, or misinformed district and state guidelines (e.g., Maki et al., 2015; McNicholas et al., 2018) may also maintain such practices but are of a different kind given their authority in clinical practice. For instance, school psychologists may continue to use the IQ-achievement discrepancy model despite decades of evidence against its use (Aaron, 1997;

Correspondence concerning this article should be addressed to:
Ryan L. Farmer, (901) 678-2313, r.farmer@okstate.edu, Oklahoma State University School of Teaching, Learning, and Educational Sciences, Stillwater, OK 74078

Dombrowski et al., 2004). To be clear, the promotion of a practice is neither problematic nor does it suggest that the practice is wholly without merit. In fact, a critical component of effective scientific communication is the promotion of scientific findings (Kappel & Holmen, 2019), which is necessary for disseminating effective practices. However, effective science communication and training have 'trust' as a central goal (see B. Cook et al., 2018 for a discussion of the role of trust in science). Lewis and Wai (2021) discussed the importance of communicating the *limits* of our understanding and the uncertainty of our claims. Hype, an exaggerated and potentially harmful kind of promotion, is the endorsement of a practice above and beyond the available evidence, or in direct contradiction to the evidence that is available at the time of the claim. When this occurs and exaggerated claims go unquestioned, LVPs may be adopted in lieu of practices whose evidentiary basis is more firmly established. Nevertheless, we stipulate that for scientific psychology to progress, there needs to be a balance of innovation and critique. As noted by Lilienfeld and colleagues (2015), "Reflexive dismissal of the new and untested is ill advised as is blind acceptance" (p. xxii). Being able to distinguish between scientific innovation and hype is critical for achieving this balance.

Healthy Self Doubt

Whether we choose to call it epistemic humility (Kidd, 2016; Lilienfeld et al., 2017), philosophic doubt (see Deitz, 1982), or simply "healthy self-doubt" (Meichenbaum & Lilienfeld, 2018, p. 23), the premise is a critical behavior for school psychologists and a pillar of EBP (National Association of School Psychologists, 2020). Adopting the attitudes consistent with healthy self-doubt means (a) acknowledging that some of the beliefs we have are fallible and that (b) review of evidence for and against those beliefs is necessary to calibrate one's position. Lilienfeld et al. (2017) suggested that helping students to develop healthy self-doubt is an underlying tenet of both the scientist-practitioner and clinical-science models of training—we extend this claim to also include school psychology programs. Graduate programs that train students to function as scientist-practitioners likely do so to improve students' awareness of their own blind spots and knowledge gaps so that they can better

evaluate their own beliefs and claims about clinical practices (Lilienfeld et al., 2017). Thus, despite the emphasis on research productivity, the ideological strengths of scientist-practitioner programs are the emphasis on scientific thinking and the development of professional skepticism. Regardless of whether a training program explicitly adheres to this model of training, we argue that critical thinking and scientific literacy are foundational to most, if not all, recognized training standards in school psychology (e.g., NASP, 2010). Recognizing that we are all susceptible to cognitive biases, such as illusory correlation and causation (Chapman & Chapman, 1969; Lassiter et al., 2002), establishing these dispositions is an important first step in beginning to understand the inherent uncertainty a clinician must confront in professional decision-making tasks. From this starting point, a school psychologist's confidence in a course of action is based on the quality and quantity of available evidence for and against a given practice as there are few, if any, practices that have been found to be infallible.

These philosophical and aspirational underpinnings are not easy to implement nor are they easy to teach. Learning to balance increased skepticism with epistemic humility is challenging. This practice is further complicated when supportive evidence is exaggerated or when contradictory evidence is devalued, minimized, or even suppressed. In particular, when school psychologists must confront multiple information sources competing for their time and attention (e.g., social media, clinical lore, trade publications; Youngstrom et al., 2015). For this reason, graduate students may benefit from explicit instruction in recognizing the warning signs of hype and overconfidence as they learn to navigate the EBP literature (Lilienfeld et al., 2017).

The Promotion of Low Value Practice

Pratkanis (1995) described a number of approaches often used to promote LVPs, such as setting unattainable but idealized goals, leading with high-probability requests, and self-anointed credibility. Whereas Pratkanis described methods used to promote LVPs, he also described methods that proponents have used to redirect criticism of their claims. For instance, it may be asserted that researchers cannot understand a practice's utility because they are not engaging

directly in that particular practice (e.g., they are not practicing school psychologists). Despite the populist notions invoked by such appeals, these approaches serve to change the conversation from evaluating the evidence for a claim or position to raising doubt about the expertise or intentions of critics, and may discourage skeptical individuals from asking questions, thereby functioning to avoid careful critical review of the quality of available research evidence. It is necessary to point out that these strategies are sometimes also used to promote practices that are regarded by the scientific community as valid clinical approaches (Wilson, 2003). The distinction appears to be in how the promoter handles and presents the available evidence. We believe it important to be aware of these strategies as there is also value in considering strategies often used in the promotion of school psychological assessment instruments and practices.

Lilienfeld and Meichenbaum (2018) provided a checklist for hype in psychotherapy that provides school psychologists with a series of "warning signs" that a particular clinical practice may be of low value. They suggested that checklists of this kind may be helpful in identifying questionable tactics often used to market practices or to exaggerate their evidentiary basis. In the interest of facilitating the critical evaluation of psychological assessment practices, we provide a provisional list of warning signs to prompt school psychologists to further evaluate assessment claims in clinical science. Again, these strategies are offered to promote a more skeptical and critical evaluation of claims rather than outright rejection of claims. To clarify, the presence or absence of these provisional indicators does not mean that the claims being made should automatically be dismissed or accepted. But the determinant of adequacy is determined by the quality and quantity (replication) of evidence.

Warning Signs of Hype in Psychological Assessment

In the following section, we provide a brief discussion of 12 different potential warning signs of hype in school psychological assessment. These strategies may not be unique to LVPs and they should not be considered an exhaustive explication of all of the tactics or practices and forms of argumentation

that have historically been associated with hype movements in psychological science.

Evasion of Peer Review

Although imperfect, peer-reviewed journals are regarded by scientists as the gold-standard for the development, exchange, and evaluation of scientific ideas. Whereas many proponents of LVPs make regular contributions to the peer-reviewed literature, all too often problematic assessment practices are disseminated and popularized in non-refereed forums (e.g., workshops, non-empirical books and chapters, and podcast discussions). This is not to suggest such forums do not contain useful information (especially as digital platforms become more popular in our profession); instead, we suggest that when these venues serve as the primary (or only) vehicle(s) for dissemination of these practices, school psychologists should view them with a higher degree of skepticism (Washburn et al., 2019). For instance, publishing mechanisms for tests and test data are largely protected from peer review until the test is already on the market, and already promoted and sold as clinically useful. While many publishing companies should be commended for sharing their data with independent researchers for evaluation afterwards, these evaluations are rarely integrated into the instrument's technical documentation and therefore less likely to be read by users.

Unqualified Belief in the Infallibility of Clinical Judgment

When LVPs have psychometric shortcomings, proponents may imply that they can be overcome through the skilled use of clinical judgment. However, this ignores the considerable errors associated with overreliance on this approach to test interpretation (e.g., Garb, 1998), including an increased risk of confirmation bias, (i.e., the inclination to seek out or interpret evidence in a manner consistent with one's prior beliefs [Nickerson, 1998]). Some scholars go so far as to describe clinical judgment as an almost mystical process. For example, when describing a popular approach to interpreting intelligence test scores (i.e., "Intelligent Testing" [IT; A. S. Kaufman, 1994]), Fletcher-Janzen (2009) noted that IT "ascends to the concrete where all deductive and inductive judgments are guided by theory, translated by the

clinician, and synthesized into an elegant whole" (p. 25). Rather than caution users against over-interpreting or misinterpreting data and guarding against potentially committing Type I (false-positive) errors, some interpretive guidebooks suggest that preference in clinical assessment should guard against committing the inverse. Such recommendations overlook the fact that abnormal scores are common in the normal population and thus inferring disorder or disability from such observations may lead to over-pathologizing healthy clients (Odland et al., 2015). As noted by Faust (2007), appeals to "integrate all of the data" (p. 67) as a safeguard against decision-making error assumes that validity is cumulative and that each unique piece of information has an additive effect on our ability to make more effective clinical decisions. While the intuitive appeal of this synthetic exercise and the desire to develop an idiographic understanding of the individual is understandable, such practices may reduce judgmental accuracy (see Bowes et al., 2020).

Lack of Connectivity and Failure to Self-correct

Well-developed theories and methods in the psychological sciences typically exhibit a degree of connectivity. That is, as new research findings emerge, aspects of the theory or method are modified accordingly. A useful analogy is the crossword puzzle, with individual answers filled in based on clues, but then re-evaluated each time another clue is answered to determine whether the initial conclusion remains a good fit (see Haack, 1993). Mutual support of the pieces is obtained when various pieces of the puzzle fit well together. However, assessment claims that contradict existing evidence (without providing compelling evidence of their own to counter these concerns) should be viewed critically. For instance, claims that results from cognitive ability tests can be used to determine treatment for students struggling with academic content are inconsistent with the empirical findings that suggest such test results do not accurately predict specific learning disabilities (McGill et al., 2018) and generally reflect poor support for aptitude-by-treatment interactions (Burns et al., 2016; Kranzler & Floyd, 2019, pp. 413-431).

Self-correction of beliefs is necessary and ongoing. LVP often lack this key characteristic and may be marked by an insularity where negative research findings are dismissed or are not acknowledged in any meaningful way. As previously mentioned, proponents of these methods may also dismiss countering claims by using tactics such as suggesting that critics do not have the requisite knowledge to critique the technique in question or that they do not understand the procedures to render a valid opinion on the matter. Additionally, proponents may also invoke elaborate post hoc explanations to explain away negative research results. While these explanations may be plausible, the fact that they were not disclosed as potential limitations a priori raises questions about their veracity (Meehl, 1967).

Potential Presence of Conflicts of Interest

Any conflict of interest, financial or otherwise, poses an increased risk of bias or motivated reasoning. Although current estimates are not available, Truscott et al. (2004) found that undisclosed financial conflicts of interests were rampant in the assessment literature in school psychology and there is no available evidence to suggest that the state of our science has improved in any meaningful way since, or that these conflicts of interest (COI) do not apply to other subfields (e.g., clinical psychology) given the shared commercial interests in play. Although a COI does not negate a person's expertise or the veracity of the content they provide, COIs may be a harbinger for undisclosed bias. However, when potential COIs related to the products and processes a person is attempting to advance are not disclosed, it raises concerns about the degree to which all available evidence is being faithfully disclosed. We invite the reader to consider how frequently they have observed presenters of commercial products (e.g., the publisher of a test) at conferences share contradictory evidence regarding their product. For example, it is not uncommon for proponents of a theory or practice to only selectively report research results that are supportive, while "file-drawering" negative evidence and creating the artificial impression that all is well (Boccaccini et al., 2017). As a consequence, school psychology practitioners and trainers are encouraged to consider the degree to which potential COIs may promote such allegiance effects. As noted by Gibbons (2015), commercial COIs often serve as an incubator that permit LVPs to flourish by impeding scientific self-correction from taking place.

Appeals to Eminence

In some cases, questionable psychological theories or practices feature a proponent who has attained authority or eminence in their respective discipline (Pratkanis, 1995). In some cases, these individuals may be respected members of the academy and/or skilled practitioners with advanced content area expertise. Eminence may be amplified by social media where critical review is often absent (Koppl, 2010). As noted by Gambrill (2012), these effects are particularly insidious when a practice is popular and widely accepted; the pressure to conform often conspires against those that raise crtical questions about the proposed method. Beyond the aforementioned commercial stakes that may incentivize and perpetuate hype movements, it is also important to consider the motivations that may undergird one's desire to maintain their leadership role regarding a popular clinical practice. For example, proponents have likely invested years of time, effort, and intellectual capital developing various assessment methods and techniques. It may be difficult, if not impossible, to overcome this psychological "sunk-cost" (Olivola, 2018) in recognizing that one's efforts have not been fruitful.

Ad Hoc Credentialing Programs

There are now several certification programs that have been developed in the field that appear to be designed primarily to promote the use of particular assessment techniques. Whereas these certification programs may serve to legitimize controversial assessment methods, some have questioned the utility of these types of credentialing pathways (e.g., Pelletier et al., 2004). Aside from the costs involved to obtain these credentials, it is unclear to what degree their development has led to improved assessment practices. Gambrill (2012) noted, "There is a huge bogus credential industry fueled by the importance attributed to credentials and the assumption that they yield competence to practice" (p. 38). In navigating this landscape, it is important to evaluate whether a particular credential or professional development program is accredited or recognized by a legitimate professional organization, follows recognized practices in conferring advanced titles in professional psychology (e.g., American Board of Professional Psychology), and charges reasonable fees for the certification being sought. Moreover, school psychologists should confirm that the skills taught by these credentialing programs lead to functional improvements for students by careful evaluation of supporting literature; they may check to see if studies have been published on the specific named practice, whether those effects have been replicated by independent researchers, and if contradictory evidence exists.

Dismissing Scientific Methods as Misinformation

Rather than debate the merits of their proposed methods, when confronted with negative research evidence, proponents of hyped practices often dismiss the methodologies employed by critics as being narrow in scope or outdated; these arguments are used even when the proponents of hype movements have relied primarily on those same methodologies to support their premises in the past and continue to use those methods to support current efforts. For instance, proponents for interpreting index scores from cognitive ability tests may dismiss interpreting scores primarily representing g by claiming it to be a statistical artifact, despite evidence for the various scores stemming largely from the same methodological practices (e.g., confirmatory factor analysis; McGill & Dombrowski, 2019). When clear methodological or analytical justification is not provided, this should be regarded as cherry-picking of preferred results (Meehl, 1978).

Appealing to Cash Validity

As noted by Frazier and Youngstrom (2007), there is a tendency among consumers to assume that more expensive tests and assessment approaches translate to better outcomes as compared to less expensive (or sometimes free) approaches. This fallacy appeals to a relationship between assessment utility and instrument (or training) cost, when such a relationship likely does not exist (Thomas & Callan, 1992). Put simply, just because an assessment instrument costs more does not mean that it is more useful than available technologies that are more cost effective. For example, despite a number of rating scales available for purchase to help school psychologists identify the function of a student's challenging behavior, there is little evidence to suggest these rating scales correspond to gold standard assessment practices for identifying functional hypotheses (Iwata et al., 2013). While

these measures are found to be superior to strictly descriptive assessment, the most well-validated (e.g., the Questions About Behavioral Function; Smith et al., 2012) are available for free online. That some free measures are likely equivalent—and perhaps superior—to paid-measures seems to be true for traditional psychological rating scales as well (Becker-Haimes et al., 2020).

Grandiose Claims Despite Stagnant Technology

Although it is frequently asserted that various assessment technologies are new or modern, most advances to our instrumentation have been incremental and not disruptive (Frazier & Youngstrom, 2007; Thiel & Masters, 2014). For example, the Wechsler scales of intelligence features many subtests that are direct descendants from measures developed during the Army testing program from WWI (Kevles, 1968). Yet, proponents of modern assessment methods suggest that advances in technology now permit school psychologists to accurately use these technologies to diagnose and select treatment for sundry conditions despite little compelling evidence to support these claims. One example is profile analysis of cognitive test results which, despite some changes, have largely remained the same over the past 20 years (McGill et al., 2018) while another is the continued use of projective testing in clinical settings (Lilienfeld et al., 2000). Appeals to novelty should not be accepted unless appropriate supporting evidence is furnished a priori (Glutting et al., 2003).

Rebranding Previous Ideas Without Addressing Research Criticism

In a comprehensive review, McGill and colleagues (2018) noted that it is remarkable how modern approaches to test interpretation share the same characteristics as original approaches that are now regarded within the canon as outdated and empirically unsupported. In this way, many contemporary approaches are simply reparameterizations of previous practices with little, if any regard for the shortcomings associated with previous versions of those approaches that are well documented. In this way, there appears to be a collective devotion within the field to these particular LVPs (e.g., cognitive profile analysis) that allows them to flourish despite little to no compelling research evidence. As noted presciently by Sagan,

"Sparse and poor popularizations of science abandon ecological niches that pseudoscience promptly fills. If it were widely understood that claims to knowledge require adequate evidence before they can be accepted, there would be no room for pseudoscience" (2011, p. 6). It is important for consumers to be able to identify when a practice is revolutionary or merely "old wine being put into new wineskins."

Neuro-Realism, Neuro-Essentialism, & Neurobabble

Three interrelated concepts—neuro-realism, neuro-essentialism, and neurobabble—play an interesting role in contemporary assessment. Neuro-realism is the notion that neurological images (e.g., functional magnetic resonance images [fMRI]) bring an objectivity or realism to the topic that is not possible with psychological or behavioral theory (Racine et al., 2005). This notion was famously explored by McCabe and Castel (2008) who found that images of brain scans increased perceptions of credibility of related scientific research. Relatedly, neuro-essentialism is a reduction of complex human characteristics to the brain, or to interactions within the brain. This type of reductionist perspective may use the brain as a shortcut for complex phenomena (e.g., "the brain can help us forget painful memories") or may suggest that related technology can evaluate the brain (Racine et al., 2005), as if it were the hard disk in a computer. Neurobabble is neuroscientific information that is irrelevant to, but presented along with, a scientific claim (Misheva, 2020), and has been shown to increase the credibility of such claims. Neurobabble may work due to the appeal of neuro-realism or -essentialism, but may boil down to providing an ostensibly more substantiated perspective of a complex or abstract phenomenon. Neurobabble is a unique variation of invoking scientism (see Haack, 2012), which uses the trappings of science (e.g., "scientific rigor," "evidence-based," "based in science") and the anticipated reader's deference to science to bypass critical appraisal. Whereas neuroscientific research has great potential for psychology, expecting immediate, direct application may be overly optimistic (e.g., Schwartz et al., 2016). Psychological assessment practices that invoke neurological jargon or explanation may be accurate, but they may also be invoking the intuitive appeal of neuro-based explanations.

School psychologists could benefit from carefully asking "What will these data tell me?" and "How will these data impact my decision making?" in an effort to redirect questions about neurological functioning to socially valid outcomes. While neurological functioning may—and likely does—mediate outcomes for students, framing assessment purpose in this way may help school psychologists to identify and scrutinize evidence for and against a given procedure. For instance, it is intuitive that tests purporting to measure neurological functioning would be useful in the diagnosis of attention deficit hyperactivity disorder (ADHD) given evidence of neurological differences among individuals with and without the disorder. However, such physical (e.g., fMRI) or cognitive (e.g., Working Memory Index) neuromarkers function as poor diagnostic indicators for ADHD (DuPaul & Stoner, 2014, pp. 34-35). The use of neurological information may be included and used appropriately by test developers and promoters, and the presence or reference to neurological information does not immediately indicate misinformation. However, we contend that such neurological data should be linked to student outcomes in order to be useful in practice.

Evidential Quality in Clinical Assessment

Carefully evaluating the underlying evidence for a claim should be the norm rather than the exception (Dombrowski et al., 2021). This leads to at least two crucial questions: How much and what type of evidence do we need? For the first question, no one can say 'how much' evidence is necessary for an individual to believe a claim; this is a debate that has received substantial attention in philosophy (e.g., Haack, 1993), but only passing treatment in psychological practice. Furthermore, the question stem, 'how much...', may imply a sort of bean counting rather than consideration of the comprehensiveness of the available evidence (whether that is a few sources or many). For the second question, how one intends to interpret and use an argument greatly affects the kinds of evidence one might need; again, there is no simple answer. For evaluating claims related to assessment, with regard to their comprehensiveness and their kind, the school psychologist must carefully evaluate how data from the assessment procedure will be interpreted and used (D. A. Cook et al., 2015;

Kane, 2013).

For example, when searching the assessment literature, it is common to encounter a set of studies that support a preferred interpretive approach for an instrument and another set of studies that suggest the opposite conclusion. As noted by Faust (2007), direct contradictions in the literature must be handled with care and selectivity in determining which evidence should be retained. It is not best handled by assuming that all information should be integrated equally. If five pieces of information of low evidential value (for instance, due to methodological gaps) support a preferred approach and one or two pieces of higher value indicate otherwise, most school psychologists would do well to give more weight to the evidence afforded by the former set.

A focused approach to the evaluation of the validity of an instrument or process is possible by identifying the types of inferences to be made using a particular approach. Kane (2013) suggests developing an interpretation/use argument which carefully defines how test data are to be used—and by extension, are useful. More simply, a school psychologist has a hypothesis about how a particular type of test data is useful, and then can go about collecting data to corroborate or falsify that hypothesis (Table 1 next page). For the researcher, this involves the collection of data from a sample of the population of interest using appropriate methodology and analysis. However, for clinicians, this likely means referring to the available evidence in the peer-reviewed literature—and short-run empiricism[1]. As a comprehensive discussion of evidence for assessment claims is beyond the scope of this manuscript—and would likely warrant multiple chapters to accomplish successfully—interested readers should consider Kane's (2013) framework and D. A. Cook et al.'s interpretation (2015).

Kane's (2013) approach to evaluating claims relies heavily on a school psychologist's understanding of measurement theory, research methodology, and logical reasoning; further buttressing the conclusions of Lilienfeld and colleagues (2017) that a scientist-practitioner, or clinical-science, model of training is essential to produce graduates who are prepared

[1]Short-run empiricism was defined by Cronbach (1975) as monitoring responses to treatment, and making adjustments as necessary (p. 126), and specifically recommended in school psychology by Reschly et al., (1997).

for these challenges. For example, Table 1 outlines a series of measurement and psychometric concepts that are often used to evaluate evidentiary claims about individual instruments and assessment practices in the evidence-based assessment literature.

Table 1

Types of Evidence by Claim Type

Claim	Types of Evidence That may Support This Claim
Text X is useful for diagnosis/classification	• Static psychometric adequacy of relevant scores • Correlation with gold standard assessments • Diagnostic accuracy • Sensitivity • Specificity • Positive & negative predictive value • Accuracy • Positive & negative likelihood ratio • ROC Curve analysis • Discrimination analysis • Longitudinal prediction of outcomes
Test X can help identify treatment	• Static psychometric adequacy of relevant scores • Longitudinal prediction of outcomes • Evidence dependent upon effectiveness of treatment
Test X is useful for progress monitoring	• Static psychometric adequacy of relevant scores • Scores are sensitive to change
Test X measures Y	• Static psychometric adequacy of relevant scores • Correlation with other measures thought to access the same construct • MTMM data • Factor analysis data • Item response theory data

Note. Table based in part on D. A. Cook et al. (2015) and Kane (2013a, 2013b, 2013c).

Surveys suggest that the statistical and measurement research methods training in contemporary training programs may not be sufficient to support critical familiarity with such a broad spectrum of approaches (e.g., Aiken et al., 2008). However, such competencies for test use have been articulated for over two decades (Turner et al., 2001). Furthermore, we acknowledge that much of the discussion around hype and various approaches to hyping assessment science are

applicable to other areas of school psychological practice, including treatment, consultation, and so forth as evinced by the debates regarding the utility of omnibus prevention models such as MTSS. As such, instruction regarding hype should not be limited to assessment or examples from assessment.

Implications for Training

Preparing students to detect hype, to evaluate assessment claims, and—generally—to think scientifically about assessment clearly spans more than a single assignment or course. Recommendations offered here are provisional and warrant evaluation and likely modification. Pedagogy regarding scientific thinking has a rich and progressive research agenda (e.g., Travers et al., 2016) but is relatively novel in school psychology. Lilienfeld et al. (2017) offered a provisional list of topics grouped by course-type that could be integrated and may help to develop graduate students' epistemic humility. Similarly, Table 2 provides a list of topics related to epistemic humility, detecting hype of assessment claims, and evaluating assessment claims and suggestions for how those topics may be included across common school psychology graduate coursework. Fortunately, a great number of the topics listed in Table 2 are already taught in school psychology programs (e.g., reliability, validity); however, others may not be taught as often (e.g., racial, ethnic, and gender biases) or may need to be integrated (e.g., how invalid tests may appear useful).

Table 2

Content to promote awareness of biases and epistemic humility by course

Content	ASMT	RM	PE	PS
Types of hype	x		x	
Types of assessment claims	x	x		
Procedures for evaluating assessment claims	x	x		
Kane's Validation Framework	x	x		
Types of reliability	x	x		
Evaluating reliability		x		
Types of validity evidence	x	x		
Evaluating validity		x		
Clinical-statistical debate	x		x	x
Biases, logical fallacies, and heuristics	x	x	x	x

	ASMT	RM	PE	PS
Clinical judgment errors & safeguards	x		x	x
Base rates & Bayes Theorem	x	x		
How invalid tests may appear useful	x		x	x
Ethnic, racial, & gender biases	x		x	x
Past and present misuses of IQ tests	x		x	
Evaluating and selecting assessment instruments	x	x	x	x
Research methods as safeguards against bias		x	x	x
Ethnic and cultural variables may establish boundary conditions		x	x	
The importance of becoming aware of one's biases	x	x	x	x
The need for ideological and intellectual diversity in psych science		x	x	x
Difference between ethnic diversity and stereotyping			x	x
Importance of informed consent	x	x	x	x
How can (and can't) science inform ethical decisions	x		x	x
Ethical dangers of clinician under- and overconfidence	x		x	x
Ethics of clinical research		x	x	
Using progress monitoring to improve outcomes	x		x	x
Disconfirming hypotheses		x		x
Considering alternative explanations	x	x	x	x

Note. Table partially based on Lilienfeld et al. (2017). ASMT = assessment; RM = research methods; PE = professional ethics; PS = practice and supervision

More work is necessary to understand how to best prepare students to engage in critical evaluation of assessment evidence. In the following section, we will review instructional strategies that may help students to detect hype and evaluate claims.

First, we recommend that trainers explicitly expose students to instruction on the risks associated with using LVPs (Lilienfeld, 2007; Lilienfeld et al., 2019), detecting hype in psychotherapy (Meichenbaum & Lilienfeld, 2018) and assessment, assessment frameworks (Hunsley & Mash, 2007), and evaluating claims (Travers, 2016). This type of explicit instruction may occur as early as an introductory course in school psychology, but may also clearly fit in coursework on assessment, measurement and statistics, research methods, and ethics. Explicit instruction may include providing definitions, giving examples from the field, or selecting exemplars from outside of school psychology that may be less controversial and easier to teach such as learning styles or whole language reading instruction. Schmaltz and Lilienfeld (2014) discuss teaching students scientific thinking and healthy skepticism by way of discussing paranormal investigators and homeopathy. Not only may this content be engaging for students, it may provide opportunities to discuss blatant examples of hype, the prevalence of hype in student's everyday lives, and strategies of identifying underlying claims. Given that multiple exemplar training is a beneficial way for students to learn how to apply new skills, beginning with extreme examples and then moving to more nuanced examples may provide students with an opportunity to learn and generalize knowledge about hype. Consider the following pedagogical example.

As Schmaltz and Lilienfeld (2014) suggest, ghost hunting television shows may be an excellent opening to discuss hype and the importance of scrutinizing claims. In such programs, the hosts often discuss their status within their community (*appeals to eminence*) and expertise with various technology (similar to *appeals to clinical judgment*), the technical qualities of their instruments (the use of science jargon is similar to *neurobabble*), and so on. An instructor might challenge students to watch a portion of such a program and report back on the types of hype they identify. They may select a low-value assessment practice and briefly review how hype is used as part of its promotion. For instance, appeals to clinical judgment are common in the interpretation of projective measures (Lilienfeld et al., 2008) and cognitive profile data (e.g., Macmann & Barnett, 1997). Similarly, some assessment proponents may use neurobabble, such as various brain training programs and their related assessments, to make their instrument seem like it's based firmly in neurology.

In addition to the previous example, there are a number of resources available online and in the peer-reviewed literature to help build scientific thinking skills, including popular science articles on LVPs (e.g., French, 2013), Cuijpers and Cristea's (2016) paper on questionable practices regarding therapy development, and critical discussions of research (e.g., https://sciencebasedmedicine.org). Other approaches might include a book club for Dan Willingham's (2012) book entitled *When Can You Trust the Experts? How to Tell Good Science from Bad in Education*;

challenging students to identify hype at state, national, and international conferences; or encouraging students to review online brochures and slides for examples of hype.

As this is a developing area of research, especially in school psychology, the authors have developed a shared repository on Open Science Framework (https://osf.io/shpez/) to collect resources. Readers who are interested are invited to contribute by e-mailing the lead author.

Detecting hype is crucial to evidence-based practice, but is not the ultimate goal. The next step is to identify the claim being made by the proponents of the instrument. For instance, in promoting instruments on social maladjustment, promoters claim that their instrument can discern when a student's challenging behavior is due to social maladjustment and when it is due to other emotional causes. While the marketing material of such scales may invoke expertise and appeal to clinical judgment, it may lack coherence with the literature (e.g., Costenbader & Buntaine, 1999). That said, regardless of whether it is hyped using one strategy or many, the goal is to determine whether it adequately discriminates between students with social maladjustment and emotional disturbance[2].

Discerning the purpose of an assessment instrument requires clarity regarding what the obtained data may say about the student, how that data may be used in decision making, and so forth (see Dombrowski et al., 2021). As a result, students may benefit from explicit instruction on assessment frameworks such as evidence-based assessment[3] (Hunsley & Mash, 2007; Youngstrom et al., 2015) and Kane's (2013) approach to validity arguments to realize the purpose of various assessments. Given knowledge of the topic, school psychologists then need to determine if the instrument must have a minimal amount of evidence or if it must be accompanied by a significant amount of evidence. This notion was eloquently

summarized by Asimov (1983) as, "I'll believe anything, no matter how wild and ridiculous, if there is evidence for it. The wilder and more ridiculous something is, however, the firmer and more solid the evidence will have to be." (p. 43). Effectively, the more grandiose the claim, the more evidence the proponent or publisher will need to provide; this is particularly true when a claim contradicts long-standing, replicated empirical outcomes. We can likely agree that the amount of evidence one needs to believe that a curriculum-based measure tells us something about a student's academic performance when used as a screening tool is decidedly less than the amount of evidence one would need to believe that a rating scale can distinguish between social maladjustment and emotional disturbance. However, as discussed previously, the 'amount' of evidence one needs is unclear and is an open debate and will depend on whether the measure is being used for low- or high-stakes decision-making. That said, Kane's (2013) approach to validation helps to address the types of evidence one might need. Instruction regarding the varied skills necessary to locate existing research evidence; to conduct novel, rigorous and transparent research; and to integrate evidence into one's practice are all challenges that graduate trainers in school psychology programs across the country already face.

Conclusion

The work that school psychologists do is of critical importance, and the implementation of evidence-based practices is paramount to avoid as much harm as possible while doing as much good as possible (see Lilienfeld et al., 2019). Adopting a sense of humility about our own practice- and theory-related beliefs is crucial to successfully navigating these responsibilities, and that doing so may help us to evaluate our beliefs and claims. However, it is not always easy to parse claims about evidence for a given assessment practice, especially when some claims are hyped. Hype tactics may seem harmless, but may also disarm critical thought or be used as evidence in their own right (Willingham, 2012). We should always be skeptical of our own beliefs and of the claims made by others; however, it is not feasible to critically evaluate every claim—no one has the time, resources, or training to accomplish such a task (Gambrill, 2008). Readers are cautioned that the presence of one or

[2]Given the absence of a documented empirical distinction between social maladjustment and emotional disturbance, one might conclude that any such instrument must meet a significant burden of proof before it could be used in practice as it contradicts the existing literature base.

[3]The terms "evidence-based assessment" and "evidence-based practice" may be invoking scientism in the use of the term "evidence-based" and we remind readers that hype alone is not an indicator of false information, and a thorough review of the underlying claims are necessary. Simply, referring to a practice as "evidence-based" does not make it so (Haack, 2012, Willingham, 2012).

many of the warning signs presented by Pratkanis (1995), Meichenbaum and Lilienfeld (2018), or those uniquely presented here do not reduce the credibility of individual scientific claims. However, because similar tactics are used to mask or suppress negative evidence or to exaggerate the effectiveness or efficacy of practices, these warning signs may serve as useful discriminative stimuli for a more comprehensive critical evaluation of underlying evidence.

Successfully preparing graduate students to be better able to spot hyped claims in assessment and better able to critically evaluate the evidence for such claims are not easy tasks, and there is much work ahead of us. Indeed, though we have provided a number of examples in this manuscript, we have not addressed all instances of hype within our field. Because these tasks are so varied and sometimes difficult to discern, it is important to consider the specific outcomes we might expect from successfully promoting these skills in graduate students. In the short term, we might observe greater acceptability of evidence-based practices, recognition of the limits of our current evidence base, and the ability to detect when claims are more or less supported by available evidence. Ideally, we would want these findings to be stable over time and, if so, should expect better clinical outcomes for students who do well with such content. In doing so, it is important to reinforce to future school psychologists that while this work may force us to confront uncomfortable truths and potentially disregard long-standing "sacred cows," these should not be regarded as moments of despair; they illustrate progress in our field through the process of scientific self-correction.

References

Aaron, P. G. (1997). The impending demise of the discrepancy formula. *Review of Educational Research*, *67*(4), 461–502. https://doi.org/10.3102/00346543067004461

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*(1), 32-50. https://doi.org/10.1037/0003-066X.63.1.32

Allen, R. A., & Hanchon, T. A. (2013). What Can We Learn from School-Based Emotional Disturbance Assessment Practices? Implications for Practice and Preparation in School Psychology. *Psychology in the Schools*, *50*(3), 290–299. https://doi.org/10.1002/pits.21671

American Psychological Association. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*(4), 271–285. https://doi.org/10.1037/0003-066X.61.4.271

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. Author. http://www.apa.org/ethics/code/principles.pdf

Asimov, I. (1983). *The roving mind*. Prometheus Books. https://www.worldcat.org/title/roving-mind/oclc/9503941

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 & Waldman (Eds.), *Psychological Science Under Scrutiny* (pp. 323–339). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119095910.ch16

Becker-Haimes, E. M., Tabachnick, A. R., Last, B. S., Stewart, R. E., Hasan-Granier, A., & Beidas, R. S. (2020). Evidence base update for brief, free, and accessible youth mental health measures. *Journal of Clinical Child & Adolescent Psychology, 49* (1). 1-17. https://doi.org/10.1080/15374416.2019.1689824

Bowes, S. M., Ammirati, R. J., Costello, T. H., Basterfield, C., & Lilienfeld, S. O. (2020). Cognitive biases, heuristics, and logical fallacies in clinical practice: A brief field guide for practicing clinicians and supervisors. *Professional Psychology, Research and Practice*, *51*(5), 435–445. https://doi.org/10.1037/pro0000309

Burns, M. K., Petersen-Brown, S., Haegele, K., Rodriguez, M., Schmitt, B., Cooper, M., ... & VanDerHeyden, A. M. (2016). Meta-analysis of academic interventions derived from neuropsychological data. *School Psychology Quarterly*, *31*(1), 28. https://doi.org/10.1037/spq0000117

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, *74*(3), 271–280. https://doi.org/10.1037/h0027592

Chronbach, L. J. (1975). The two disciplines of scientific psychology. *American Psychologist, 30*(2), 116-127. https://doi.org/10.1037/h0076829

Cook, B. G., Loyd, J. W., Mellor, D., Nosek, B. A., & Therrien, W. J. (2018). Promoting open science to increase the trustworthiness of evidence in special education. *Exceptional Children*, *85*(1), 104–118. https://doi.org/10.1177/0014402918793138

Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education*, *49*(6), 560–575. https://doi.org/10.1111/medu.12678

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. https://doi.org/10.1176/appi.ps.60.5.671

Cuijpers, P., & Cristea, I. A. (2016). How to prove that your therapy is effective, even when it is not: a guideline. *Epidemiology and Psychiatric Sciences*, *25*(5), 428–435. https://doi.org/10.1017/S2045796015000864

Deitz, S. M. (1982). Defining applied behavior analysis: An historical analogy. *The Behavior Analyst*, *5*(1), 53–64. https://doi.org/10.1007/BF03393140

Dombrowski, S. C., Kamphaus, R. W., & Reynolds, C. R. (2004). After the Demise of the Discrepancy: Proposed Learning Disabilities Diagnostic Criteria. *Professional Psychology: Research and Practice*, *35*(4), 364–372. https://doi.org/10.1037/0735-7028.35.4.364

Dombrowski, S. C., McGill, R. J., Farmer, R. L., Kranzler, J. H., & Canivez, G. L. (2021). Beyond the rhetoric of evidence-based assessment: A framework for critical thinking in clinical practice. *School Psychology Review,* Advance online publication. https://doi.org/10.1080/2372966X.2021.1960126

DuPaul, G. J. & Stoner, G. (2014). *ADHD in the schools: Assessment and intervention strategies* (3rd ed.). Guilford Press.

Farmer, R. L., McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2020, December). Why questionable assessment practices remain popular in school psychology: Instructional materials as pedagogic zehicles. *Canadian Journal of School Psychology*. Advanced online publication. https://doi.org/10.1177/0829573520978111

Faust, D. (2007). Decision research can improve can increase the accuracy of clinical judgement and thereby improve patient care. In S. O. Lilienfeld & W. T. O'Donohue (Eds.), *The great ideas of clinical science: 17 principles that every mental health professional should understand* (pp. 49-76). Routledge.

Fletcher-Janzen, E. (2009). Intelligence testing: Bridging the gap between classical and romantic science in assessment. In J. C. Kaufman (Ed.), *Intelligent testing: Integrating psychological theory and clinical practice* (pp. 15–29). Cambridge University Press.

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence*, *35*(2), 169–182. https://doi.org/10.1016/j.intell.2006.07.002

French, C. (2013, April 27). The unseen force that drives Ouija boards and fake bomb detectors. *The Guardian*. http://www.theguardian.com/science/2013/apr/27/ouija-boards-dowsing-rods-bomb-detectors

Gambrill, E. (2008). Evidence-based (informed) macro practice: Process and philosophy. *Journal of Evidence-Based Social Work*, *5*(3–4), 423–452. https://doi.org/10.1080/15433710802083971

Gambrill, E. (2012). *Propaganda in the helping professions*. Oxford University Press.

Garb, H. N. (1998). *Studying the clinician: Judgement research and psychological assessment*. American Psychological Association.

Gibbons, P. (2015). *The science of successful organizational change: How leaders set strategy, change behavior, and create an agile culture*. Ft Pr.

Glutting, J. J., Watkins, M. W., & Youngstrom, E. A. (2003). Multifactored 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* (pp. 343–374). The Guilford Press.

Gross, T. J., Farmer, R. L., & Ochs, S. E. (2019). Evidence-based assessment: Best practices, customary practices, and recommendations for field-based assessment. *Contemporary School Psychology*, *23*(3), 304–326. https://doi.org/10.1007/s40688-018-0186-x

Haack, S. (1993). *Evidence and inquiry: Towards reconstruction in epistemology*. Blackwell.

Haack, S. (2012). Six signs of scientism. *Logos & Episteme, 3* mo (1), 75-95. https://doi.org/10.5840/logos-episteme20123151

Hunsley, J., Lee, C. M., Wood, J. M., & Taylor, W. (2015). Controversial and questionable assessment techniques. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology, 2nd ed* (pp. 42–82). The Guilford Press.

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

Iwata, B. A., Deleon, I. G., & Roscoe, E. M. (2013). Reliability and validity of the functional analysis screening tool. *Journal of Applied Behavior Analysis*, *46*(1), 271–284. https://doi.org/10.1002/jaba.31

Kane, M. T. (2013). The argument-based approach to validation. *School Psychology Review*, *42*(4), 448–457. https://doi.org/10.1080/02796015.2013.12087465

Kappel, K., & Holmen, S. J. (2019). Why science communication, and does it work? A taxonomy of science communication aims and a survey of the empirical evidence. *Frontiers in Communication*, *4*. https://doi.org/10.3389/fcomm.2019.00055

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. Wiley & Sons.

Kevles, D. J. (1968). Testing the Army's intelligence: Psychologists and the military in World War I. *The Journal of American History*, *55*(3), 565-581. https://doi.org/10.2307/1891014

Kidd, I. J. (2016). Intellectual humility, confidence, and argumentation. *Topoi*, *35*(2), 395–402. https://doi.org/10.1007/s11245-015-9324-5

Koppl, R. (2010). The social construction of expertise. *Society*, *47*(3), 220–226. https://doi.org/10.1007/s12115-010-9313-7

Kranzler, J. H., & Floyd, R. G. (2019). Remediating student learning problems: Aptitude-by-treatment interaction versus skill-by-treatment interaction. In M. K. Burns (Ed.), *Introduction to school psychology: Controversies and current practices.* (pp. 413-431). Oxford University Press.

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

Lassiter, G. D., Geers, A. L., Munhall, P. J., Ploutz-Snyder, R. J., & Breitenbecher, D. L. (2002). Illusory causation: Why it occurs. *Psychological Science*, *13*(4), 299–305. https://doi.org/10.1111/j.0956-7976.2002..x

Lewis, N. J., & Wai, J. (2021, February). Communicating what we know, and what isn't so: Science communication in psychology. *Perspectives on Psychological Science*. Advance online publication. https://doi.org/10.1177/1745691620964062

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2*(1), 53-70. https://journals.sagepub.com/doi/10.1111/j.1745-6916.2007.00029.x

Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (Eds.). (2015). *Science and pseudoscience in clinical psychology* (2nd ed.). Guilford.

Lilienfeld, S. O., Lynn, S. J., O'Donohue, W. T., & Latzman, R. D. (2017). Epistemic humility: An overarching educational philosophy for clinical psychology programs. *Clinical Psychologist*, *70*(2), 6–14.

Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., & Latzman, R. D. (2019). The Insufficiently appreciated raison d'être of evidence-based practice. In S. Dimidjian (Ed.), *Evidence-based practice in action: Bridging clinical science and intervention*. Guilford Press.

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1* (2), 27-66. https://doi.org/10.1111%2F1529-1006.002

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2008). Why questionable psychological tests remain popular. *The Scientific Review of Alternative Medicine, 10,* 6-15.

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

Maki, K. E., Floyd, R. G., & Roberson, T. (2015). State learning disability eligibility criteria: A comprehensive review. *School Psychology Quarterly*, *30*(4), 457–469.

https://doi.org/10.1037/spq0000109

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, *107*(1), 343–352. https://doi.org/10.1016/j.cognition.2007.07.017

McFall, R. M. (1991). Manifesto for a science of clinical psychology. *The Clinical Psychologist*, *44*(6), 75–88.

McFall, R. M. (1996). Making psychology incorruptible. *Applied and Preventive Psychology, 5*(1), 9-15. https://doi.org/10.1016/S0962-1849(96)80021-7

McFall, R. M. (2000). Elaborate reflections on a simple manifesto. *Applied and Preventive Psychology*, *9*(1), 5–21. https://doi.org/10.1016/S0962-1849(05)80035-6

McGill, R. J., & Dombrowski, S. C. (2019). Critically reflecting on the origins, evolution, and impact of the Cattell-Horn-Carroll (CHC) Model. *Applied Measurement in Education*, *32*(3), 216–231. https://doi.org/10.1080/08957347.2019.1619561

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

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

McNicholas, P. J., Floyd, R. G., Woods Jr., I. L., Singh, L. J., Manguno, M. S., & Maki, K. E. (2018). State special education criteria for identifying intellectual disability: A review following revised diagnostic criteria and Rosa's Law. *School Psychology Quarterly*, *33*(1), 75–82. https://doi.org/10.1037/spq0000208

Meehl. P. E. (1967). Theory-testing in psychology and physics. A methodological paradox. *Philosophy of Science, 34*(2), 103-115. https://doi.org/10.1086/288135

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

Misheva, E. (2020). Neuromyths, neurobabble and pseudoscience: The complex relationship between the neuro-disciplines and education. In E. Misheva (Ed.), *Child neuropsychology in practice: Perspectives from educational psychologists* (pp. 9–27). Springer International Publishing. https://doi.org/10.1007/978-3-030-64930-2_2

National Association of School Psychologists. (2020). *The professional standards of the National Association of School Psychologists.* National Association of School Psychologists. https://www.nasponline.org/standards-and-certification/nasp-practice-model/about-the-nasp-practice-model

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology: Journal of Division 1, of the American Psychological Association*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Odland, A. P., Lammy, A. B., Perle, J. G., Martin, P. K., & Grote, C. L. (2015). Reaffirming normal: The high risk of pathologizing healthy adults when interpreting the MMPI-2-RF. *The Clinical Neuropsychologist, 29*(1), 38-52. https://doi.org/10.1080/13854046.2015.1005675

Olivola, C. Y. (2018). The interpersonal sunk-cost effect. *Psychological Science, 29*(7), 1072-1083. https://doi.org/10.1177/0956797617752641

Pelletier, S., Gibbons, P., & Shapiro, M. (2004). The application of neuropsychology in the schools should not be called school neuropsychology: A rejoinder to Crespi and Cooke. *The School Psychologist*, *58*(1), 17–24.

Pigilucci, M. (2018). Scientism and pseudoscience: in defense of demarcation projects. In M. Pigilucci (Ed.), *Science unlimited: The challenges of scientism*. University of Chicago Press.

Pratkanis, A. R. (1995). How to sell a pseudoscience. *The Skeptical Inquirer*, *19*, 19–25.

Racine, E., Bar-Ilan, O., & Illes, J. (2005). FMRI in the public eye. *Nature Reviews Neuroscience*, *6*(2), 159–164. https://doi.org/10.1038/nrn1609

Reschly, D. J. & Ysseldyke, J. E. (1997). School psychology paradigm shift. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (3rd ed.). 17-31. National Association of School Psychologists.

Sagan, C. (2011). *The demon-haunted world: Science as a candle in the dark*. Random House Publishing Group.

Schwartz, S. J., Lilienfeld, S. O., Meca, A., & Sauvigné, K. C. (2016). The role of neuroscience within psychology: A call for inclusiveness over exclusiveness. *American Psychologist*, *71*(1), 52–70. https://doi.org/10.1037/a0039678

Silva, M. R., Collier-Meek, M. A., Codding, R. S., Kleinert, W. L., & Feinberg, A. (2020). Data collection and analysis in Response-to-Intervention: A survey of school psychologists. *Contemporary School Psychology, 25*, 554-571. https://doi.org/10.1007/s40688-020-00280-2

Smith, C. M., Smith, R. G., Dracobly, J. D., & Pace, A. P. (2012). Multiple-respondent anecdotal assessments: an analysis of interrater agreement and correspondence with analogue assessment outcomes. *Journal of Applied Behavior Analysis*, *45*(4), 779–795. https://doi.org/10.1901/jaba.2012.45-779

Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools, 52*(10)*, 1031-1045*. https://doi.org/10.1002/pits.21802

Thiel, P., & Masters, B. (2014). *Zero to one: Notes on startups, or how to build the future* (Illustrated edition). Currency.

Thomas, J. M., & Callan, S. J. (1992). An analysis of production cost inefficiency. *Review of Industrial Organization*, *7*(2), 203–225. https://doi.org/10.1007/BF00158141

Travers, J. C. (2017). Evaluating claims to avoid pseudoscientific and unproven practices in special education. *Intervention in School and Clinic*, *52*(4), 195–203. https://doi.org/10.1177/1053451216659466

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. https://doi.org/10.1521/scpq.19.2.166.33311

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*(12)*, 1099–1113*. https://doi.org/10.1037/0003-066X.56.12.1099

Washburn, J. J., Lilienfeld, S. O., Rosen, G. M., Gaudiano, B. A., Davison, G. C., Hollon, S. D., Otto, M. W., Penberthy, J. K., Sher, K. J., Teachman, B. A., Peris, T., & Weinand, J. (2019). Reaffirming the scientific foundations of psychological practice: Recommendations of the Emory meeting on continuing education. *Professional Psychology: Research and Practice*, *50*(2), 77–86. https://doi.org/10.1037/pro0000235

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

Wilson, N. (2003). Commercializing mental health issues: Entertainment, advertising, and psychological advice. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 425-459). Guilford.

Youngstrom, E. A., Choukas-Bradley, S., Calhoun, C. D., & Jensen-Doss, A. (2015). Clinical Guide to the Evidence-Based Assessment Approach to Diagnosis and Treatment. *Cognitive and Behavioral Practice*, *22*(1), 20–35. https://doi.org/10.1016/j.cbpra.2013.12.005