

Psychometric Evaluation Competency: The International Psychometric Evaluation Certification (IPEC) in the Era of Evidenced Based Assessment Within the Forensic Environment

Scott Whitmer

Whitmer & Associates

Cynthia P. Grimley

Cynthia P. Grimley & Associates

Abstract. *This article will draw upon relevant clinical, psychometric, forensic methodology, and empirical data to illustrate the utility of a certification for competency in psychometric evaluation. The International Psychometric Evaluation Certification (IPEC) was conceptualized by the American Board of Vocational Experts (ABVE) in 2014 after two years of research and membership collaboration. The IPEC mission is to endorse and support qualitative and quantitative empirically based methods, ethically driven standards, efficiency of evaluation, competency of experts/evaluators, and legally defensible work. Evidence Based Assessment (EBA) criteria share these common goals that underlie the IPEC mission and will serve to align with the EBA national mandate. In the tradition of collaboration and transparency, the ABVE wishes to call upon all evaluators from all corners of the evaluation industry to come forward to share in the vision that will not only improve methodology, competency, ethics, efficiency, and legally defensible work but align with EBA goals. This article will also serve to highlight methodology that is emerging within the discipline of Forensic Vocational Evaluation that will build upon research and broaden methodology associated with psychometric theories, methods and applications. Our discipline and related specialties wish to build upon this important element of empirical work to demonstrate to future legislative leaders that regardless of differences in mission and vision within the counseling and expert communities, we as psychometric evaluators stand united in endorsing a methodology, standards and purpose that will continue to effectively serve the courts and general public.*

The eventual demarcation of philosophy from science was made possible by the notion that philosophy's core was "theory of knowledge," a theory distinct from the sciences because it was their foundation... Without this idea of a "theory of knowledge," it is hard to imagine what "philosophy" could have been in the age of modern science.

— Richard Rorty, in *Philosophy and the Mirror of Nature* (1979, p. 132).

The above quotation by Richard Rorty (1979) illustrates an important point in our quest as evaluators to understand where our roots lie most deeply. The theory of knowledge (epistemology) and the nature of being (ontology) indeed lie at the foundation of philosophy. Philosophy is at the foundation of psychology and therefore we are rooted in the study of knowledge and measurement of being. Forensic vocational expert work is rooted in theory and application of psychology and psychometric evaluation and hence a social science. Psychological measurement is a critical and contemporary scientific method to accurately understand human behavior, cognition, affect, perception and conscious awareness that is

included in our assessments. In regard to ontological exploration of psychology, notwithstanding its psychometric methods, researchers believe that a Kuhnian "scientific revolution" preparadigmatic state is at hand. The paradigm shift is moving to a unifying psychological theoretical assumption that exists on a continuum of theoretical perspectives that include situational realism (Empiricism), developmental evolutionary psychology (Piagetian) and the Tree of Knowledge (ToK) unified theory (matter, life, mind and culture corresponding to four classes of science: physical, biological, psychological and social). The scientific revolution exists within the dialectic of metaphysical and practical experimental pre-

dictions (Marsh & Boag, 2014). This pre-paradigmatic position claims that the three aforementioned theories have an empiricist thread that will serve to unite psychology (and its sub-specialties in counseling) and to provide a meta-theoretical framework to explain the level of complexity that further frames the ontological and epistemological foundations of psychology and measurement. This is important for the history and future of psychometric testing because the foundation of evaluation of psychological human measurement rests on such theoretical framework, differing schools of thought and various methodologies. We must be aware not only of the weaknesses and strengths of psychometric testing but also the weaknesses and strengths of our theoretical foundations so that we do not dogmatically attach to a single outdated paradigm of scientific inquiry, thereby impeaching the science or our measurement methodologies. On the pragmatic and reality end of the continuum, the Association of Test Publishers (Harris, 2006), makes it known that the *Standards for Educational and Psychological Testing* (hereafter called the *Standards*) are critical not only for evaluators but also publishers in building high quality, legally defensible, valid and reliable tests. Ultimately publishers state that too many standards divorced from the reality of test construction make it difficult to keep the tests aligned with such *Standards*.

In our quest for valid and reliable measurement we acknowledge that social science deals with the measurement of human cognition, behavior, affect, perception, and spirituality. Such constructs cannot always be quantifiable in perfectly predictable patterns because humans do not follow a prescribed or predictable set of behaviors all of the time. Therefore, social science relies upon statistically quantifiable and empirically based methods to define what is more probable within a standardized and norm referenced sample. In the discipline of evaluation and psychometrics, we also rely upon qualitative methods to provide incremental validity to psychometric measurement or in other words use of other relevant and technical data that support one's hypothesis or predicted outcome. We also utilize triangulation of data to validate measurements both quantitatively and qualitatively (Leech & Onwuegbuzie, 2007). Incremental validity and triangulation are methods that when combined with psychometric measurement of some psychological construct or set of constructs, add validity because of the different multiple sources that add credibility to the question at hand (Hunsely, 2003). A psychometric instrument or test that provides a strong coefficient correlation between a construct and a predicted behavior or outcome may indeed be worthless to the evaluator and stakeholders without sound qualitative methods of inquiry such as a structured interview, questionnaire, projective test, or consultative interaction. In our work as evaluators, counselors and forensic experts, we typically work within the confines of the case study method or N

of 1 sample (Field & Choppa, 2005). It is posited that since our sample size is N of 1, the psychometric measurement becomes exponentially important and predictive when we can rely upon a reliable and validated instrument that allows us to infer and generalize with accuracy. We can with a certain statistical confidence validate that an examinee has a reading level at the 88th percentile within his peer age group or a mechanical aptitude at the stanine score of 7 within his educational peer group. In essence we can ascertain a confidence range and a known error rate to relay to the court that a person has achieved a level of academic achievement, language, cognitive ability, relevant personality, intelligence, aptitude or any construct that is measurable. If performed ethically and with standardization protocol, psychometric measurement can create the parameters of a case for which all other methodological work is based. If performed with haste, unfair selection of instruments, administrator error or any number of other competency related issues, the forensic evaluator will not only provide invalid psychometric data, but a number of other evaluation methods within the assessment that rely upon the psychometric data will be tainted (Kaplan & Saccuzzo, 2013). With the potential for a forensic assessment to go so wrong and the potential to harm the examinee, it is with the purpose and mission of the IPEC to create standards through competency criteria, ethics standards, and empirical research so that the potential for quality, efficiency, efficacy, reliability and validity are expanded and achieved in the industry. Psychometric measurement does enjoy a scientific reputation that our courts of law recognize and rely upon quite regularly (Kaplan & Saccuzzo, 2013). Furthermore, the intent of the IPEC is to strengthen this trust relationship with the courts and general public from a national forensic platform through the IPEC mission to develop a validated instrument to measure competency (Grimley & Whitmer, 2014; Goodwin & Leech, 2003).

A quantitative hypothesis contains a null proposition and an alternative proposition that is either validated or not through statistical analysis (Kaplan & Saccuzzo, 2013). Once a hypothesis is generated, researchers and evaluators can set out to use a number of methods to test the hypothesis to determine validity and reliability. The qualitative method investigates the why and how of decision making. The quantitative method uses a traditional randomized controlled study to affirm the hypothesis or refute it with correlative or predictive power. Both qualitative and quantitative methods are invaluable for measuring human constructs (Groth-Marnat, 2009). Thomas Samuel Kuhn, American physicist and philosopher, popularized the term "paradigm shift" in reference to scientific knowledge (Hergenhahn & Henley, 2014) in "The Structure of Scientific Revolutions." Kuhn believed that scientific fields undergo periodic paradigm shifts that result in a thesis-antithesis tug of war. Eventually one theory combines with another, and a synthesis to the dialectic resolves the

competing positions by supporting both positions in proving a more valid and reliable outcome of measurement. We have seen this occur numerous times in the field of psychology and the vocational rehabilitation industry alike. In the vocational evaluation field we have accepted the synthesis of qualitative and quantitative measures at the research level, evaluation level and at the legal pragmatic level.

There is little doubt left in the social scientific community that both clinical judgment and psychometric testing provide incremental validity, efficiency and outcome efficacy. It is widely accepted that psychometric testing without methodological clinical inquiry is all but useless and probably harmful (APA, 2010) as illustrated in the *APA Ethical Principles of Psychologists and Code of Conduct; Assessment 9.06*. Relative to a qualitative perspective of measurement, quantitative (objective) psychometric testing typically improves upon what the clinician has drawn out in his/her clinical methods. However, congruency may not be met if the qualitative data do not agree with the quantitative data, at which point the evaluator must query a new hypothesis, get a new referral question or question his own methodology (Kaplan & Saccuzzo, 2013; Groth-Marnat, 2009). The triangulation of qualitative and quantitative data indeed is a method that we undertake several times in the evaluation of N-1 throughout the span of assessment (Robinson, 2014).

Foundational Consideration

Vocational Author/Expert, Mary Barros-Bailey (Robinson, 2014) provided a poignant depiction of the history and future of forensic vocational evaluation within the larger context of vocational consulting across many legal venues. She was right in stating we as a discipline must look to outside disciplines to help us understand and define the directions that will provide relevant answers to the question of how to ensure our place within the legal community and also in history. Mary Barros-Bailey posits there are many other forensic sciences that have not been in existence as long as forensic vocational consultation (FVC) yet our discipline has fewer credits in well-known vocational and career public management venues. This finding is a call for our future leaders to think more broadly about our future growth (Robinson, 2014) to ensure we are woven deeper into the legal and legislative fabric of evaluation and overall health care system.

In the early development of the IPEC many sister disciplines (clinical and forensic) have been cited and researched to build upon the framework for standards and excellence. A few related disciplines researched to help the ABVE and the IPEC to verify that directions in foundation building include social work, medical, mental health, life care planning, nurse case management, school psychology, psychology, substance abuse, and criminal evaluation professionals. A rele-

vant example of related disciplines is the American Psychological Association (APA) Specialty Guidelines for Forensic Psychology (APA, 2013) an ethical guidepost for forensic psychologists who wish to rely upon established ethical forensic practices. The ABVE and the IPEC will be looking to relevant forensic guidelines from organizations such as the APA and from other disciplines to develop its own special guidelines for its unique members.

Evaluation Competency of Members

The need for evaluation competency and measurement of competency was originally conceptualized in the forensic vocational community through the American Board of Vocational Experts (ABVE, 2014). The ABVE Board of Directors acknowledged the need to build upon the scientific standards on which we base our work and to ensure competency among members for future decades to come. Initially, there was a call from members to revive the Certified Vocational Evaluation (CVE). However, after two years of attempted collaboration with the Board of Certified Vocational Evaluators (BCVE), it became apparent they did not share the same mission and vision as those of the ABVE Board of Directors, with regard to competency, a broader concept of psychometric evaluation and EBA criteria. As the IPEC concept began to gain momentum, it became apparent that many members and non-members were interested in building a methodologically based certification that encompasses not just vocational evaluation methods, but broader psychometric evaluation methods. Psychometric evaluation methods are known to emanate from many psychological theories, methods and applications (AERA, APA & NCME, 1999) that are being used in the forensic and rehabilitation industry today. Through formal and informal surveying, it was discovered that many ABVE members are quite competent in selecting, administering, analyzing, scoring, interpreting and synthesizing data that are gleaned from psychological measures, however many ABVE members are not, and wish to obtain competency. Moreover, the ABVE wanted to look to other disciplines to determine if professional competency in psychometric evaluation was an issue. Research revealed that sister disciplines have also struggled with evaluation competency. As the literature demonstrates, many master's level graduates lack competency in a broad use of psychometric instruments (Baker & Ritchey, 2009; Brodenhorn & Skaggs, 2005). In many instances those graduate level evaluators who were academically trained to administer, score and interpret psychological measures did not exercise their qualifications or did not achieve a broader competency across testing domains due to the nature of work they were performing in rehabilitation. Moreover, many business models around the nation relied upon one or two professionals within their ranks to perform the work as Evaluator (Baker & Ritchey, 2009). As a result, many master's

level experts have the academic preparation, understanding of theory, standardized administration protocol, scoring, analysis, synthesis, and report writing abilities but lack a full current and relevant competency to exercise said skills. The ABVE Board of Directors further posits that sister disciplines at the masters level of counseling and rehabilitation have the same need for standards building to create competency in the counseling and forensic settings.

Evidenced-Based Assessment

As made salient in David Barlow's (2005) article "What's New About Evidence-Based Assessment?" the Agency for Healthcare, Research and Quality has a mission to improve quality, safety, efficiency, and effectiveness of health care for all Americans. Barlow points out that the final report of the President's New Freedom Commission on Mental Health commits to a private-public partnership in guiding evidence-based assessment/evidence-based practice (EBA/EBP) and to expand it to the work-force in providing evidence-based practice. Other researchers note seven interrelated steps in the operational definition of EBP including commitment to apply EBP, translating needs into treatment questions, obtaining and critically appraising evidence, applying results, evaluating outcomes, examining the role of the client (examinee), and emphasizing the mindful and systematic evaluation of intervention to clients (Baker & Ritchey, 2009; Hunsley & Mash, 2005). Such EBP steps can be easily transferred to EBA outcomes and measures. Other research investigating EBA practices and psychometric test use among master's level clinicians versus doctoral level clinicians (Jensen-Doss & Hawley, 2010) found that master's level clinicians held attitudes towards psychometric instruments that were based on concerns of practicality and their benefit over clinical judgment alone. Jensen-Doss & Hawley (2010) essentially explained that such attitudes from master's level clinicians were related to lack of training. Researchers suggest that front line clinicians/evaluators receive training to overcome misconceptions and attitudes about psychometric test use to meet EBA goals.

The meaning and intent of Evidence-Based Assessment (EBA) is rooted in the national movement to practice and assess based on evidence outcome effectiveness. What drives evidenced based outcomes is managed health care goals to obtain empirically based, effective, efficient, safe, cost effective and quality care services (Barlow, 2005). EBA/EBP began with the focus on treatment and then moved into assessment, realizing that it was just as important to link treatment outcomes to what works, as it was to determine if assessment methodology was adding to what works (Hunsley & Mash, 2005). Regardless, EBA is a world health care phenomenon that will continue to grow in its effort to ensure that the most current science is congruent with effective out-

comes, consistent with treatment quality, and equating cost effectiveness to all other measures within the EBA model. The IPEC is viewed as an industry beacon that will align the discipline of assessment within the parameters of EBA while at the same time honoring empirically based methods that rely upon quantitative and qualitative evaluation.

In the process of defining The Council in Rehabilitation Education (CORE) accreditation standards for Research and Program Evaluation courses (Schultz & O'Brien, 2008), researchers made salient the relevance of EBP. EBP's roots take hold from the Hippocratic Oath in regard to principles of beneficence and the potential to inflict harm to patients. The Commission on Rehabilitation Counselor Certification (CRCC) embodies many of the EBP tenants that include methods grounded in relevant theory, empirical foundation, and efficacy of treatment or intervention. EBP is also qualified by efficiency or what works, methodologies that have a scientific foundation, and implications that qualified counselors will be a critical part of the EBP formula. Further, Schultz & O'Brien (2008) posit that evidence for effectiveness should not only include randomized controlled trials but also mixed methods research, quasi-experiments with equating, regression discontinuity designs and single case (N-1) designs. They further elaborate there is a dearth of EPA and EBP literature within the rehabilitation counseling discipline. The absence of research with regard to EBP practice standards is likely because of the diversity of rehabilitation research (subspecialties) and lack of experimentally controlled research within the field.

Research in the clinical psychology discipline suggests there are 12 steps that can be implemented in EBA and research ranging from identifying most common diagnoses to soliciting and integrating patient preferences (Youngstrom, 2013), many of which already exist in the field of vocational evaluation. EBA or EB medicine in this case, suggests that much of what is compiled in the prediction, prescription and process can be reformulated to measure what is most important to the patient and clinical outcome. Youngstrom (2013) posits, as do these authors, that much of what we need to align with EBA is available; we just need to be creative, organized and mindful to implement efficiency, efficacy, scientific rigor and client outcome preference. It is hoped that the IPEC and its EBA focus will generate the interest and motivation of many practitioners and researchers alike to write, research and apply psychometric methods that seem to bind together our industry through theory and application. Researchers in the disability and rehabilitation industries (Leahy & Arokiasamy, 2010; Tarvydas, Addy, & Fleming, 2010) acknowledge EBP/EBA and testing standards will effect and inform practice and policy on a national level.

Legal Competencies and Psychometric Methods

There are five primary aspects of forensic evaluation and testifying that define most of what we do as forensic evaluators with the acknowledgement that there are many more subparts to each that are not mentioned here:

1. The clinical/qualitative interview.
2. Psychometric Testing/Evaluation.
3. Analysis-synthesis of medical/vocational/historical/claims facts.
4. Analysis and reporting of the findings/opinions and recommendations.
5. Testifying in court.

Psychometric Evaluation of the examinee is an integral part of forensic evaluation (Robinson, 2014; AERA/APA/NCME, 1999) and without psychometric evaluation, experts in many instances will not be able to make scientifically based inferences upon which experts and the court so importantly rely. In the role of forensic testing there have been six factors identified that make up a model of legal competencies and they include (Heilbrun, 1992): 1. Functional abilities relevant to legal competency; 2. The context in which competency must be demonstrated; 3. Causal inference between observed deficits and legal ability; 4. Interaction between ability and specific demands of situation; 5. Judgment by decision maker to determine person-situation incongruence to warrant a finding of incompetence; 6. Disposition of legal response to decision maker's finding. While the six legal competencies appear to be most relevant to criminal cases, they can apply to civil cases. Heilbrun (1992) points out the Trier of Fact has the right to consider political, moral, and environmental values in the final influence of his/her decision in reference to civil and criminal cases. In the case where testing is known or suspected to have a racial or cultural bias even with the evidence of rigorous empirical structure, the courts may depend more upon the racial, cultural or political factors that may confound the test results. Forensic evaluators need to be aware of the legal context in which the tests will be viewed and be able to provide an explanation of the biases, potential errors and results that the lay person can understand.

Psychometric evaluation is the one aspect of our work about which we can say with empirical degree of certainty, that the theory from which we apply our findings has a known error rate (*Daubert v. Merrell Dow Pharmaceuticals*, 1993). Much of the rest of our work is process and product that can be summed up in the *Daubert* rubric of tested technique/theory, subjected to peer review or publication and theory or technique generally accepted in the industry. At a closer look, psychometric evaluation and methods may be refuted under all four of the *Daubert* criteria if the psychometric evaluation cannot be upheld. With regard to psychometric evaluation, it is hypothesized that if the theory and technique

that is relied upon have no error rate, have unknown normative characteristics, or lack sufficient strength, then the remaining three criteria under *Daubert* open the door for inquiry and probable legal challenge that may result in impeachment of the expert and his or her opinions to the remaining portions of the assessment (Wise, 2006). Federal Rules of Evidence based on the *Daubert* ruling is cited below.

The US Supreme Court (1993) case of *Daubert v. Merrell Dow Pharmaceuticals* ruled under *Federal Rules of Evidence* (FRE Rule 702) that there are four primary considerations or questions that are relevant for admissibility of scientific testimony (Field & Stein, 2002; Sackett, 2011):

1. Has a theory or technique been tested?
2. Has a theory or technique been subjected to peer review and or publication?
3. Does a theory or technique have a known error rate or standards?
4. Has a theory or technique been generally accepted in the industry?

In our pursuit to provide objective data through empirical means we must remember what are considered empirically supported measures. Much can go wrong resulting in false positives, administrator error, interpretive errors, cultural error, and a number of other errors and biases that can spoil the data (Feuer, 2011; Wise, 2006). It is further noted that regardless of empirical validity and rigorous methodology, the Trier of Fact may choose to rely upon ecological, political, cultural or clinical explanations of the case for final decision making. The potential for this disconnect between the court and empirical findings should prepare Forensic Evaluators to ensure psychometric testing results make sense in the light of the overall story of the case and its fact pattern (Borum, Otto, & Wiener, 2000). We must be careful not to divorce psychometric findings from other parts of the assessment that might not have relevant and reasonable explanations and may be as equally damaging as unattended administration errors and biases of testing that taint the validity of psychometric findings. In addition we must be careful not to confuse standards for expert testimony with the standards on psychometric testing (Sackett, 2011) because while they converge to support validity and methodology, each has its divergent and differing tenants for standard setting. Courts do take a more practical and less theoretical view on validity and tend to focus on evidence of test content and consequences while the Standards emphasize sound test development and ethical testing practices (Sireci & Parker, 2006).

Psychometric Empirical Assumptions

Cohen and Swerdlik as cited by Robinson and Drew (2014) identified seven basic assumptions with regard to testing and assessment:

- (1) Psychological traits and states exist;
- (2) Psychological traits and states can be quantified;
- (3) Test-related behavior predicts non-test-related behavior;
- (4) Tests and measurement techniques have strengths and weaknesses;
- (5) Various sources of error are part of the assessment process;
- (6) Testing and assessment can be conducted in a fair and unbiased manner;
- (7) Testing and assessment benefits society” (p. 132).

The authors contend that all seven of these testing assumptions can benefit society, in addition to three more that could be added: (8) Testing and assessment exist within a qualitative clinical context; (9) must be done with practiced and known standards and ethics; (10) should attempt to conform to EBA principles. The three additional assumptions help us tie together what the mission of the IPEC has in mind for the forensic vocational evaluator and related counseling disciplines. Within the entire assessment process of vocational rehabilitation and forensic vocational evaluation, research in our discipline looks to discover or perhaps just acknowledge validating methods that can be EBA supported, are considered sound qualitative empirical methodology and are currently being practiced in single case, N-1 studies. Such methodologies that are currently used in the evaluation disciplines include incremental validity and triangulation validity.

A thorough definition of validity that connects scientific rigor to assessment outcomes is taken from Downing’s (2003) explanation of meaningful interpretation of assessment data:

Validity refers to the impartial, scientific collection of data, from multiple sources, to provide more or less support for validity hypothesis and relates to logical arguments, based on theory and data, which are formed to assign meaningful interpretations to assessment data.

Incremental Validity

Incremental validity is indeed a method that is used in forensic vocational assessment and other assessment disciplines. Incremental validity in part is defined by the assumption that other verifiable sources enhance the validity of an empirical instrument’s psychometric characteristics. We see incremental validity in use by way of multiple uses of qualitative measures, multiple use of psychometric instruments, and multiple use of informants (Hunsely, 2003). Incremental validity has a research definition but also an assessment and evaluation meaning. Essentially, incremental validity in evaluation of a single case study would rely upon the method

in the following way: The evaluator may use a structured interview, a functional capacity checklist inventory and a psychometrically validated instrument to measure pain. Each evaluation method has its merits and weaknesses. The structured interview allows the evaluator to observe behavior and determine if the reported behavior is consistent with records. The inventory may reveal that the examinee is consistent or inconsistent with the structured interview on a set of known physical or mental barriers. And thirdly, the validated empirical instrument may affirm the first two measures or not, with a known error rate that can either be corroborated or refuted. The evaluator then can proceed with the remainder of the assessment with confidence that the degree of methodology of measurement is more comprehensive, than if just one technique, method or instrument were used. Hunsely (2003) calls for more research and literature analysis in the assessment disciplines and by doing so, we can further affirm the utility of incremental validity and enhance the scientific community regarding the applied value of assessment.

Triangulation Validity

Triangulation validity may or may not use a psychometric test in verifying across or within data sources to verify accuracy. Khagram and Thomas (2010), in their article “Toward a Platinum Standard for Evidence-Based Assessment by 2020,” would include two gold standards to define the platinum standard for EBA in the twenty-first century. Gold Standard I is derived from experimental methods, counterfactuals and average causal effects while Gold Standard II derives from case studies, comparative methods, triangulation and causal mechanisms. It is proposed that both Gold Standards (GS) will be prominent in defining the Platinum Standard and will underpin EBA standards. Both GS I & II have strengths and weaknesses but if combined and organized well in pursuing EBA, will ask pertinent questions such as: What is the primary focus of the assessment? What ontological premises are assumed? Finding out what works is one of the primary premises of EBA. Triangulation and comparison (counterfactual assessment) is one method within the Platinum Standard that evaluators will identify as an effective tool to define what works and is focused on outcomes rather than outputs. Triangulation and comparison assume that not just the experiment or psychometric evaluation identify the measureables, but objective measureables are defined within the context of clinical judgment and stakeholder engagement (Khagram & Thomas, 2010).

Researchers define triangulation (Guba, as cited by Leech and Onwuegbuzie, 2007) as a means of improving rigor or analysis and methodology by assessing the integrity of the inferences that one draws from more than one source. Triangulation can involve the use of multiple theories, methods,

and sources for the purpose of representation and legitimation. Representation refers to the ability to extract accurate meaning from underlying data. Legitimation refers to trustworthiness, credibility, dependability, confirmability and transferability of inferences made.

Choppa, Johnson and Neulicht (2014) make salient six forms of triangulation methods or subparts of methods that are relevant to Forensic Vocational Evaluators' (FVE) process of case conceptualization and clinical judgment. Triangulation methods 2, 3 and 5 are given operational terms by these authors for purposes of delineation and elaborated on based on Guba (1985), Leech and Onwuegbuzie, (2007) and Khagram and Thomas' (2010) conceptualizations of triangulation. The six forms of triangulation are listed in Table 1, with examples provided by the authors.

Blaikie (1991) stated that triangulation is consistent with ontological and epistemological assumptions made in qualitative methodology and that we should be clear in our use of the definition and to what methodologies we are applying it. In conclusion we do not want to misconceive quantitative methodologies with qualitative methodologies as each has different meanings and implications. It is recommended that further research on methods of Triangulation be undertaken

in the discipline of forensic evaluation and psychometric evaluation.

Ethically Based Evaluation Methods

The ABVE wishes to join the membership organizations, credentialing bodies, government agencies, test publishers and academic institutions that uphold and follow the Standards of Educational and Psychological Testing (AERA, APA & NCME, 1999). The American Psychological Association, the American Education Research Association and the National Council on Measurement in Education collaborated in 1974 to publish the Standards of Educational and Psychological Testing (AERA, APA & NCME, 1999). In 1985 and then again 1999 significant revisions were made followed by more updated editions to include the sixth edition published in 2011. The term "Standards" will be used throughout this paper in reference to the Standards of Educational and Psychological Testing (AERA, APA & NCME, 1999). It is noteworthy that a number of other well-known credentialing bodies were involved in revising the *Standards* of Educational and Psychological Testing to include Commission of Rehabilitation Counselor Certification (CRCC), National Organization for Competency Assurance (NOCA), National Board of Certified Counselors (NBCC), Society for Human Resource Management (SHRM), numerous Ameri-

Table 1

Six Forms of Triangulation Relevant to FVEs

Investigator Triangulation

Example: Analysis of multiple medical opinions that triangulate functional capacity to corroborate or refute its transferability of inference

Examinee/Client-Evaluator Triangulation

Example: Analysis of examinee subjective account of pain experience that is relevant in comparison to the medical evaluator or other non-medical evaluator account of measured response

Collateral Informant Triangulation

Example: Analysis of data collected from employers, employees, professional organizations, labor market data, family members, etc. obtained through casework and fieldwork

Theoretical Triangulation

Example: Analysis of one or more validated instruments that is founded on the same theory such as personality trait theory to match trait characteristics to occupational best-fit; or theory of vocational interest using two qualitative methods and a quantitative method to verify interest

Peer Review Triangulation

Example: Analysis of opinions, methods, technique amongst FVE peers to confirm technique, method, error rate or general acceptance

Data Triangulation

Example: Analysis of earnings records from Social Security Administration, Employment Security and pay stubs to arrive at actual earnings; Analysis of USDOL wage data, professional industry data, and local wage data to project earning capacity

can Psychological Association divisions, American Counseling Association (ACA) and Association of Test Publishers (ATP) to name a few. The *Standards* have basic tenants and guidelines that are widely acceptable across numerous counseling disciplines that set standards in selection, administering, scoring, interpreting, analyzing and synthesizing psychometric tests (Eignor, 2013). The *Standards* guide evaluators in psychometric use that is consistent with the APA code of ethics and with legal consideration for forensic evaluators. The *Standards* make transparent the intent to promote the sound and ethical use of tests for the purpose to guide valid, reliable, efficacious, fair, efficient and quality psychometric methods. Some basic principles from the Standards of Educational and Psychological Testing (APA, AERA, NCME, 2011) are selected and paraphrased below:

- The *Standards* are to provide criteria for the evaluation of tests, testing practices, and effect of test use although the test selection, use and appropriateness relying upon professional judgment.
- The *Standards* advocates that within feasible limits, the relevant technical information be made available so that those involved in policy debate be fully informed but do not dictate public policy regarding the use of tests.
- Assessment is a broader term, commonly referring to a process that integrates test information with information from other sources such as social, educational, employment and psychological data.
- Evaluation is a more narrow term that may describe a subset of tests or part of an assessment but the *Standards* are still applied.
- The *Standards* applies most directly to standardized measures generally recognized as tests or instruments that measure ability, aptitude, achievement, attitudes, interests, personality, cognitive function, and mental health.
- It is useful to distinguish between devices that lay claim to concepts and techniques of the field of educational and psychological testing and those that represent non-standardized or less standardized aids.
- When tests are at issue in legal proceedings requiring expert witness testimony it is essential that professional clinical judgment be based on the accepted corpus of knowledge in determining the relevance of particular standards in a given situation. The intent of the Standards is to offer guidance for such judgments.
- The *Standards* are concerned with a field that is evolving and therefore monitoring and revision of the *Standards* is expected in the field.
- Prescription of the use of a specific technical method is not the intent of the *Standards*, therefore alternative acceptable methods and statistics may be applied.

The Standards of Educational and Psychological Testing (AERA, APA & NCME, 1999) constitutes fifteen chapters that cover the following: Validity, Reliability and Errors of Measurement, Test Development and Revision, Scales, Norms, and Score Comparability, Test Administration, Scoring and Reporting, Supporting Documentation for Tests, Fairness in Testing and Test Use, The Rights and Responsibilities of Test Takers, Testing Individuals of Diverse Linguistic Backgrounds, Testing Individuals and Disabilities, The Responsibilities of Test Users, Psychological Testing and Assessment, Educational Testing and Assessment, Testing In Employment and Credentialing, and Testing in Program Evaluation and Public Policy.

There are many segments that are relevant to the Forensic Evaluation discipline found in Standards of Educational and Psychological Testing (AERA, APA & NCME, 1999) but a specific section on page four, the Introduction addresses testing with regard to expert witness testimony. An excerpt is illustrated below from the *Standards* to illustrate the relevance of psychometric standards for forensic evaluators:

When tests are at issue in legal proceedings requiring expert witness testimony it is essential that professional clinical judgment be based on the accepted corpus of knowledge in determining the relevance of particular standards in a given situation. The intent of the Standards is to offer guidance for such judgments.

Research shows that many master's level graduates lack competency in a broad use of psychometric instruments (Baker & Ritchey, 2009; Brodenhorn & Skaggs, 2005). In many instances those graduate level evaluators who were academically trained to administer, score and interpret psychological measures did not exercise their skills or did not achieve a broader competency across testing domains due to the nature of work they were performing in rehabilitation. Further if professional associations and graduate training programs are not familiar with the *Standards* with regard to testing, problems are likely going to be substantially higher (Camara & Lane, 2006).

In review of the National Organization For Competency Assurance Guide to Understanding Credentialing Concepts (NOCA/NCCA, 2005) governed and credentialed by the National Commission for Certifying Agencies (NCCA), it is apparent that such a credentialing body exists to lead the global effort to bring competency, standards, and responsible governance to the general public, consumers and the members of multiple disciplines who utilize psychometric evaluation. The ABVE and the IPEC will be undertaking steps to explore accreditation to ensure IPEC credentialing that involves standards set forth by the certification as well as the development of the IPEC exam.

Conclusion

It is time for the International Psychometric Evaluation Certification to establish ethical standards, competency standards, training guidelines, and a validated certification exam to take its rightful place in the psychometric evaluation community. Vocational visionaries have provided a blueprint for psychometric evaluation and research by looking to related disciplines to pursue growth in standards, methodology and research. Evidence-based assessment is systemic in our health care system and it will be critical in the forensic vocational evaluation industry and related disciplines, as we move forward with this certification. Membership of the ABVE has voted overwhelmingly to adopt the IPEC for the primary purpose of competency among members and associated disciplines. The definition of legally defensible work is congruent with EBA tenants and the IPEC goals and mission are aligned with legal competencies and EBA tenants. Furthermore, literature demonstrates that judges are expanding the pool of forensic evaluators/experts to include mental health professionals among psychologists and psychiatrists for court determination of competency (Siegel, 2008). Based on the Trier of Fact's view of who can and should be experts in the court, there is a call for training and competency for master's level providers, particularly in the psychometric evaluation portion of forensic evaluation. The gravity and magnitude of harming another human being looms large in the pursuit of psychometric evaluation. The adoption of the *Standards* (AERA, APA & NCME, 1999) is paramount in all underlying principles and development of the IPEC. The accreditation of the IPEC through a respected credentialing organization such as NOCA/NCCA will help ensure legal and ethically defensible psychometric evaluation work.

References

- American Board of Vocational Experts, Inc. (2014, March and 2014, June). American Board of Directors Board Meeting, March 27, 2014 and June 19, 2014. [Meeting Minutes].
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct*. Washington, DC: American Psychological Association.
- American Psychological Association. (2012). Guidelines for assessment of and intervention with persons with disabilities. *American Psychologist*, 67(1), 43–62. doi:10.1037/a0025892
- American Psychological Association. (2013). Specialty guidelines for forensic psychology. (2013). *American Psychologist*, 68(1), 7–19. doi: 10.1037/a0029889
- Baker, L. R., & Ritchey, F. J. (2009). Assessing practitioner's knowledge of evaluation: Initial psychometrics of the Practice Evaluation Knowledge Scale. *Journal of Evidence-Based Social Work*, 6(4), 376–389. doi:10.1080/15433710902911097
- Barlow, D. H. (2005). What's new about evidence-based assessment? *Psychological Assessment*, 17(3), 308–311. doi: 10.1037/1040-3590.17.3.308
- Blaikie, N. W. H. (1991). A critique of the use of triangulation in social research. *Quality & Quantity*, 25(2), 115. doi: 10.1007/BF00145701
- Bodenhorn, N., & Skaggs, G. (2005). Development of the School Counselor Self-Efficacy Scale. *Measurement & Evaluation in Counseling & Development*, 38(1), 14–28.
- Borum, R., Otto, R., & Wiener, R. L. (2000). Advances in forensic assessment and treatment: An overview and introduction to the special issue. *Law & Human Behavior: The Newsmagazine of the Social Sciences*, 24(1), 1–7.
- Camara, W. J., & Lane, S. (2006). A historical perspective and current views on the standards for educational and psychological testing. *Educational Measurement: Issues & Practice*, 25(3), 35–41. doi:10.1111/j.1745-3992.2006.00066.x
- Choppa, A. J., Johnson, C. B., & Neulicht, A. T. (2014). Case conceptualization: Achieving opinion validity through the lens of clinical judgment. In R. Robinson (Ed.), *Foundations of forensic vocational rehabilitation* (pp. 261–278). New York, NY: Springer.
- Daubert v. Merrell Dow Pharmaceuticals* (No. 113 S. Ct. 2786, 1993).
- Downing, S. M. (2003). Validity: On the meaningful interpretation of assessment data. *Medical Education*, 37(9), 830. doi:10.1046/j.1365-2923.2003.01594.x
- Eignor, D. R. (2013). The standards for educational and psychological testing. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA Handbook of Testing and Assessment in Psychology, Vol 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 245–250). Washington, DC: American Psychological Association. doi:10.1037/14047-013
- Feuer, M. J. (2011). Politics, economics, and testing: Some reflections. *Mid-Western Educational Researcher*, 24(1), 25–29.
- Field, T., & Choppa, A. J. (Eds.). (2005) *Admissible testimony*. Athens, GA: Elliott & Fitzpatrick.
- Field, T., & Stein, D. (2002). *Science vs. non-science and related issues of admissibility of testimony by rehabilitation consultants*. Athens, GA: Elliott & Fitzpatrick.
- Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement

- courses. *Measurement & Evaluation in Counseling & Development*, 36(3), 181–191.
- Grimley, C. P., & Whitmer, S. (2014). The Evolution of The International Psychometric Certification (IPEC). *Journal of Forensic Vocational Analysis*, 15(2), 7–12.
- Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). Hoboken, NJ: John Wiley.
- Harris, W. G. (2006). The challenges of meeting the standards: A perspective from the test publishing community. *Educational Measurement: Issues & Practice*, 25(3), 42–45. doi:10.1111/j.1745-3992.2006.00067.x
- Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior*, 16(3), 257–272. doi:10.1007/BF01044769
- Hergenhahn, B. R., & Henley, T. (2014). *An introduction to the history of psychology* (7th ed.). Belmont, CA: Wadsworth.
- Hunsley, J. (2003). Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment*, 15(4), 443–445. doi:10.1037/1040-3590.15.4.443
- Hunsley, J., & Mash, E. J. (2005). Introduction to the special section on developing guidelines for the evidence-based assessment (EBA) of adult disorders. *Psychological Assessment*, 17(3), 251–255. doi:10.1037/1040-3590.17.3.251
- Jensen-Doss, A., & Hawley, K. M. (2010). Understanding barriers to evidence-based assessment: Clinician attitudes toward standardized assessment tools. *Journal of Clinical Child & Adolescent Psychology*, 39(6), 885–896. doi:10.1080/15374416.2010.517169
- Kaplan, R. M., & Saccuzzo, D. P. (2013). *Psychological testing: Principles, applications, and issues* (8th ed.). Belmont, CA: Cengage.
- Khagram, S., & Thomas, C. W. (2010). Toward a platinum standard for evidence-based assessment by 2020. *Public Administration Review*, 70, S100–S106. doi:10.1111/j.1540-6210.2010.02251.x
- Leahy, M. J., & Arokiasamy, C. V. (2010). Prologue: Evidence-based practice research and knowledge translation in rehabilitation counseling. *Rehabilitation Education*, 24(3), 173–175. doi: 10.1891/088970110805029796
- Leech, N. L., & Onwuegbuzie, A. J. (2007). An array of qualitative data analysis tools: A call for data analysis triangulation. *School Psychology Quarterly*, 22(4), 557–584. doi: 10.1037/1045-3830.22.4.557
- Marsh, T., & Boag, S. (2014). Unifying psychology: Shared ontology and the continuum of practical assumptions. *Review of General Psychology*, 18(1), 49–59. doi:10.1037/a0036880
- National Organization for Competency Assurance. (2005). *The NOCA guide to understanding credentialing concepts*. Retrieved from <http://www.cvacert.org/documents/CredentialingConcepts-NOCA.pdf>
- Robinson, R. (2014). *Foundations of forensic vocational rehabilitation*. New York, NY: Springer.
- Robinson, R., & Drew, J. L. (2014). Psychometric assessment in forensic vocational rehabilitation. In R. Robinson (Ed.), *Foundations of forensic vocational rehabilitation* (pp. 87–131). New York, NY: Springer.
- Rorty, R. (1979). *Philosophy and the mirror of nature*. Princeton, NJ: Princeton University Press.
- Sackett, P. R. (2011). The uniform guidelines is not a scientific document: Implications for expert testimony. *Industrial & Organizational Psychology*, 4(4), 545–546. doi:10.1111/j.1754-9434.2011.01389.x
- Schultz, J. C., & O'Brien, M. (2008). Meeting CORE standards in the research and program evaluation course. *Rehabilitation Education*, 22(3), 287–294. doi: 10.1891/088970108805059372
- Siegel, D. M. (2008). The growing admissibility of expert testimony by clinical social workers on competence to stand trial. *Social Work*, 53(2), 153–163. doi: 10.1093/sw/53.2.153
- Sireci, S. G., & Parker, P. (2006). Validity on trial: Psychometric and legal conceptualizations of validity. *Educational Measurement: Issues and Practice*, 25(3), 27–34. doi: 10.1111/j.1745-3992.2006.00065.x
- Tarvydas, V., Addy, A., & Fleming, A. (2010). Reconciling evidence-based research practice with rehabilitation philosophy, ethics and practice: From dichotomy to dialectic. *Rehabilitation Education*, 24(3), 191–204. doi: 10.1891/088970110805029831
- Wise, L. L. (2006). Encouraging and supporting compliance with standards for educational tests. *Educational Measurement: Issues & Practice*, 25(3), 51–53. doi:10.1111/j.1745-3992.2006.00069.x
- Youngstrom, E. A. (2013). Future directions in psychological assessment: Combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical Child & Adolescent Psychology*, 42(1), 139–159. doi:10.1080/15374416.2012.736358

Author Notes

Scott Whitmer, Director-at-Large of the American Board of Vocational Experts, Scott Whitmer & Associates, LLC. Cynthia P. Grimley, President of the American Board of Vocational Experts, Cynthia P. Grimley & Associates, LLC. Correspondence regarding this article should be addressed to Scott Whitmer, 205 N. 4th Avenue, Yakima, WA 98902; Email: scott@whitmerandassociates.com