

© 2019, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/amp0000586

## The Enhanced Examination for Professional Practice in Psychology: A Viable Approach?

Jennifer L. Callahan<sup>1</sup>  
University of North Texas

Deborah J. Bell  
University of Missouri-Columbia

Joanne Davila  
Stony Brook University

Sheri L. Johnson  
University of California, Berkeley

Timothy J. Strauman  
Duke University

Cindy M. Yee  
University of California, Los Angeles

in press, *American Psychologist*

Author note: Jennifer L. Callahan, Department of Psychology, University of North Texas; Deborah J. Bell, Department of Psychological Sciences, University of Missouri-Columbia; Joanne Davila, Department of Psychology, Stony Brook University; Sheri L. Johnson, Sheri L. Johnson, Department of Psychology, University of California, Berkeley; Timothy J. Strauman, Department of Psychology & Neuroscience, Duke University; Cindy Yee-Bradbury, Department of Psychology, University of California, Los Angeles.

Address correspondence to: Jennifer L. Callahan, Ph.D. ABPP; Department of Psychology; University of North Texas; 1155 Union Circle #311280; Denton, TX 76205. E-mail: jennifer.callahan@unt.edu

---

<sup>1</sup> Subsequent to the first author, authors are listed alphabetically and contributed equally to the manuscript.

### Abstract

Health disciplines have increasingly required competency-based evaluations as a licensure prerequisite. In keeping with this trend, the Association of State and Provincial Psychology Boards (ASPPB) has begun to develop a second part to the Examination for Professional Practice in Psychology (EPPP). The resulting two-part examination is collectively referred to as the Enhanced EPPP. Part 1 of the Enhanced EPPP, which consists of the current exam, is designed to be an assessment of knowledge. Part 2 of the Enhanced EPPP is newly developed and intended to address the need for a competency-based evaluation. To date, ASPPB has addressed some standard facets of validity for the EPPP Part 2, but not others. In addition, the EPPP Part 2 has yet to be subjected to a broader validation process, in which the suitability of the test for its intended purpose is evaluated. Implementation of the EPPP Part 2 before validation could have negative consequences for those seeking to enter the profession and for the general public (e.g., potential restriction of diversity in the psychology workforce). For jurisdictions implementing the EPPP Part 2, failure to gather and report the evidence required for use of a test in a forensic context may also open the door for legal challenges. We end with suggestions for feasible research that could significantly enhance the validation process for the EPPP Part 2 and offer jurisdictions concrete suggestions of features to look for in determining whether and when to implement the Enhanced EPPP.

*Keywords: Licensure, Validity, Validation, Examination for Professional Practice in Psychology (EPPP), Association of State and Provincial Psychology Boards (ASPPB)*

*Public Significance Statement:* The national licensing exam for psychologists acts as a gatekeeping evaluation intended to protect public welfare. To date, the suitability and value added of the EPPP Part 2 is unclear, and ASPPB has described only limited plans to conduct

validation of the EPPP Part 2. Validation of the planned revision to the exam is of crucial significance to the entire profession and serves to protect the discipline's reputation as a health service provider.

**The Enhanced Examination for Professional Practice in Psychology: A Viable Approach?**

Almost two decades ago, the Institute of Medicine (IoM) Board on Health Care Services' Committee on the Health Professions Education Summit (2003) recommended that competency-based education become the standard of training for all health service provider disciplines. Across the ensuing years, health service psychology has made important strides in assimilating this recommendation into doctoral psychology education and training (APA, 2018; Callahan & Watkins, 2018a; Callahan & Watkins, 2018b). While there is no consensus across higher education on a standard operational definition of competency-based education (Gervais, 2016), the general approach is one that shifts curricular attention away from future-oriented goals and objectives to evaluation of present tense, realized outcomes (see O'Connell & Moomaw, 1975 for a review of conference discussions that led to that seminal distinction). Correspondingly, in keeping with IoM's recommendation, recent years have seen a shift in doctoral psychology accreditation requirements away from a focus on delineating program-level goals and objectives in future tense language (Guidelines and Principles for Accreditation of Programs in Professional Psychology; APA, 2008) towards a focus on assessment and documentation of student-centered, present-tense competency attainments (Standards of Accreditation for Health Service Psychology; APA, 2018).

A natural extension of the IoM's recommendation for competency-based education is a growing trend across health care fields towards competency-based evaluation of licensure candidates seeking to enter their profession. All major health care professions in the United States (U.S.) have developed, validated, and disseminated assessment procedures for evaluating the preparation and appropriateness of candidates for licensure. Historically, prior to the IoM recommendation, such evaluations focused on assessment of foundational knowledge using

standardized multiple-choice examinations (e.g., the Examination for Professional Practice in Psychology; EPPP). In contrast, more contemporary competency-based evaluations assess how adept a licensure candidate is with applying the requisite professional knowledge, skills, values, attitudes, and behaviors under authentic practice conditions.

Consistent with these national trends across health services, the Association of State and Provincial Psychology Boards (ASPPB) has begun to revise the existing EPPP in order to incorporate competency assessment into the evaluation of candidates for licensure as a psychologist. The primary revision of the EPPP (which ASPPB now calls the Enhanced EPPP) involves developing an additional competency-based test (referred to as Part 2) to supplement the existing, traditional foundational knowledge test (hereafter referred to as Part 1). Descriptions of Part 1's development and validity, as well as longstanding concerns associated with the exam, have already been identified and debated in the peer reviewed literature (e.g., Erikson Cornish & Smith, 2009; DeMers, 2009; DiLillo & Tremblay, 2009; Rosen, Reaves, & Hill, 1989; Ryan & Chan, 1999) and will not be repeated herein. Rather, this paper primarily focuses on the emergence of Part 2. To facilitate clarity across the larger literature, we will note where an issue pertains to both Part 1 and Part 2. As ASPPB has suggested including both Part 1 and Part 2 in the Enhanced EPPP, we will also discuss consequent issues associated with incremental validity in this paper.

### **Is Part 2 Ready for Implementation?**

Test development necessitates a systematic and coordinated approach that examines validity as well as validation before implementation. Despite the phonemic similarity between “validity” and “validation,” the two terms are associated with different meanings – accuracy versus appropriateness, respectively – and hold different implications as they pertain to test

development (Cizek, 2016; Hughes, 2018). The highly cited *Standards for Educational and Psychological Testing* (referred to simply as *Standards* hereafter; American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA & NCME], 2014) defines validity as “a unitary concept” (p. 14). Fundamental to that definition is that all validity is construct validity,<sup>2</sup> and careful adherence to the *Standards* is widely regarded as best practice for ensuring test validity (e.g., Wise & Plake, 2016).

According to public statements issued by ASPPB, the *Standards* form the basis of the technical work being conducted to develop Part 2 of the Enhanced EPPP. As Cizek (2012) notes, however, a narrow and technical focus on validity cannot determine whether a test ultimately should be used for the proposed purpose. Test validity is only one part of a larger iterative validation process (see Figure 1). The goal of such a validation process is to determine whether a test is appropriate for use (e.g., Cizek, 2012; Hubley & Zumbo, 2011; Kane, 2016; Newton & Shaw, 2013; Shepard, 2016; Sireci, 2016; Zumbo & Hubley, 2016). During validation, “It is the interpretations of test scores for proposed uses that are evaluated, *not the test itself.*” (AERA, APA, NCME, 2014, p. 11, emphasis added). Standard practice dictates that the test developer and publisher hold joint responsibility for the technical question of whether a test can be used (relying on psychometric evidence of construct validity), while stakeholders are responsible for resolving the ethical question of whether a test should be used (in light of findings from validation studies). In this paper, we seek to promote a transparent and balanced validation process that brings together the test developers and stakeholders in a productive partnership. In the sections that follow, we consider the current state of validation of Part 2.

---

<sup>2</sup> As such, types of validity (e.g., predictive validity, content validity) are not specifically delineated or considered in the *Standards*.

First, we review the facets of validity that have and have not been established. We contend that beyond the initial validity tests, ASPPB has not yet acknowledged the need for a broader, comprehensive process of validation. We highlight some concrete examples of ways in which premature use of Part 2 might be inappropriate and yield unintended consequences that could have been identified prospectively and addressed via a more thorough validation process. Premature adoption and its potential for adverse consequences also raises the question of whether Part 2 can withstand challenges in the courts, and so we pay particular attention to legal standards regarding appropriateness of test implementation.

Second, we draw attention to the present lack of involvement and buy-in from relevant stakeholders and argue for the merits of broader collaboration and the peer-review process. To provide a comparison to the test development process associated with the Enhanced EPPP, we summarize the recent focus in medical competency evaluations on careful and iterative collection of validity data and the broader process of test validation. We also provide some historical context for our expectations by briefly summarizing validation studies of the original EPPP (i.e., Part 1) in which ASPPB, jurisdictions, and independent investigators collaborated successfully.

We wish to emphasize that we are not questioning the need to assess applied knowledge or skills as part of licensure readiness, or questioning the importance of what ASPPB has undertaken. Rather, we highlight standard goals in test development that we believe have not been fully achieved in the development of Part 2 in the Enhanced EPPP, even as it is evident that other disciplines have been able to do so effectively. Without further test development and validation, it is unclear whether Part 2 is positioned to accomplish its stated aim. We end by urging the discipline as a whole to consider issues of validation and to exercise its collective

responsibility to determine whether or not there is sufficient evidence to justify and ethically implement Part 2 of the Enhanced EPPP.

### **Summary of Validity and Validation Evidence<sup>3</sup>**

Table 1 (adapted from Hughes, 2018) highlights ten forms of evidence<sup>4</sup> that may be used to establish the accuracy (validity) and appropriateness (validation) of a test such as the Enhanced EPPP. In addition to reporting the prevalence of each type of evidence in recent test development articles, Table 1 includes (a) an example of how that evidence type applies to the Enhanced EPPP, and (b) whether ASPPB has indicated plans to provide such evidence.

**Validity evidence.** As shown in Table 1, the strength of the Enhanced EPPP Part 2 stems from the careful test development process of identifying critical content via a profession-wide job tasks survey, writing items to fit that content, and evaluating item performance via item response theory (IRT) analyses. ASPPB has provided evidence for this type of validity.

The second type of evidence in the table, structural, is a relative weakness in ASPPB's communications and actual analyses. Ninety percent of test development articles report statistics associated with structural evidence of validity (Hughes, 2018), yet such analyses have not been described or reported by ASPPB to date. According to ASPPB, Part 2 of the Enhanced EPPP is intended to capture six different forms of competency: scientific orientation; assessment and intervention; relational competence; professionalism; ethical practice; collaboration, consultation, and supervision. Structural analyses, such as factor analyses or multidimensional IRT, could assess whether items cohere into unique subscales capturing these six domains (as

---

<sup>3</sup> We do not have a full list of ASPPB plans, and this review focuses on the statements offered in their description of the EPPP Part 2 (ASPPB, October, 2017).

<sup>4</sup> Table 1 validity and validation types are drawn from a review of all newly developed scales published in two of the most highly regarded psychometrics journals, *Psychological Assessment* and *Assessment*, between April 2015 and June 2016 (Hughes, 2018). We do not attempt to include all possible approaches, but focus on the most commonly used approaches. For example, we do not cover response process validity, which is an important and underutilized approach.



compared to forming one unified factor that captures a more general knowledge of the field, or the ability to use logic and intelligence to solve test items, as two examples).

Another apparent limitation in ASPPB's communication and analytic approach pertains to stability across groups (the 3<sup>rd</sup> entry in Table 1). Reflecting their appreciation for the risk of differential item functioning across subgroups of examinees, ASPPB describes efforts to include diverse individuals among job task survey respondents and Part 2 item writers, to train item writers to avoid cultural and linguistic bias, and to have an evaluation committee conduct item level screening for possible bias (ASPPB, 2017). These efforts are laudable, though perhaps overly narrow in focus (e.g., socioeconomic status and/or implicit bias do not appear to have been considered). Despite this seeming awareness, ASPPB does not describe a plan to evaluate whether or not those efforts actually produce stable validity indices, for content or structure, across subgroups.

Examination of measurement and structural invariance in organizational research provides a way to examine these issues, providing an evidence-based method to assess comparability across gender, race, ethnicity, and SES (Vandenberg & Lance, 2000). In a letter to members of the California Board of Psychology (n.d., but distributed around 1/29/2019), ASPPB's Senior Director of Examination Services (M. Turner) and the ASPPB Implementation Task Force Chair (E. Rodolfa) indicate that scores from individuals in early adopter jurisdictions will be used to set the pass point, pass rates, and determine item level psychometric data. Although beta-testing could be carried out on known groups to allow for examination of measurement and structural invariance, the letter to the California Board indicates exclusive reliance on convenience sampling that is dependent upon a cohort of applicants from early

adopter jurisdictions. We view this testing and sampling strategy as inadequate with respect to issues of both validity and validation.

**Validation evidence.** At a broad level, the remaining 8 indices found in Table 1 all pertain to validation of a test *with regard to its intended use*. ASPPB has argued that assessing the appropriateness of the Enhanced EPPP is beyond the scope of their duty, with no plans or intention to conduct any such investigations (p. 2; Communication from ASPPB, 1/29/2019). As test developers who hold responsibility for establishing technical validity, one might reasonably agree with their position that such investigations are beyond their requisite duty. It is undeniable, however, that validation to determine appropriateness is a critical component of the development process and it appears to have been neglected thus far.

Although ASPPB's mission is "[t]o support member jurisdictions in fulfilling their responsibility of public protection," we would argue that, indeed, they are not positioned to successfully complete a validation process independently. Test justification decisions must be guided by the values of the profession and the ethics of the field (Figure 1). As such, validation must be inclusive of the varied stakeholders across health service psychology education, training, and licensure, as well as the public. Validation is a major undertaking and the process might seem overwhelming. Yet, psychological science as a discipline has a huge investment and is well positioned to address the challenge. Test validation is a core facet of psychological research that is a central career goal for many psychological scientists. Accordingly, the field is rich with individuals who could contribute to this process. While this process may be outside the scope of ASPPB's mission, ASPPB is well positioned to facilitate a spirit of cooperative, stakeholder-driven validation processes that services the ultimate aim of protecting the public and builds confidence among stakeholders in the ultimate value and appropriateness of the proposed test.

Historically, during the development of the original EPPP, ASPPB encouraged and facilitated a series of cooperative validation processes (see Shrader, 1980 for a review).

### **Consequences of Implementation Without Comprehensive Validation**

As a high-stakes examination, implementation of the Enhanced EPPP before completing a comprehensive, psychometrically rigorous process of validation may have important implications for individuals seeking licensure as well as for the general public. Although not a complete listing, some prominent concerns include: diversity constriction, consequences for doctoral training, jurisdictional inconsistency, personal finance implications, and legal challenges. We discuss these concerns in the context of both the current EPPP, as well as the enhanced EPPP (which includes both Parts 1 and 2) because implementation of the Enhanced EPPP is directly linked to present use of the current EPPP.

**Failing to address diversity and inclusion issues.** Racial and ethnic minorities, men, and individuals with disabilities are all under-represented in psychology's emerging workforce (Callahan et al., 2018). Failure to develop an equitable exam may directly contribute to further diversity constriction in the future workforce and undermine the likelihood of successfully meeting the mental health needs of an increasingly diverse population (e.g., Agency for Healthcare Research and Quality, 2017; Ibaraki & Hall, 2014; Owen, Tao, Imel, Wampold, & Rodolfa, 2014; Tao, Owen, Pace, & Imel, 2015). Thus, ensuring that the EPPP Part 2 is unbiased is of crucial importance given the under-representation of minorities within psychology.

Evidence from investigators working with the second APA Task Force of the Commission on Ethnic Minority Recruitment, Retention and Training suggests that the EPPP Part 1 systematically constricts racial and ethnic diversity in the workforce (APA Commission on Ethnic Minority Recruitment, Retention, and Training in Psychology Task Force, 2008;

Bowman & Ameen, 2018). Relatedly, Sharpless and Barber (2013) found that increased student body diversity was associated with lower EPPP pass rates at the program level. Further, in a recent study that drew upon the Freedom of Information Act to gain access to the full population of data associated with a large, populous state ( $N$  examinees = 4,892), the failure rate at first exam attempt clearly varied by race and ethnicity (African/American or Black = 38.50%; Hispanic/Latinx = 35.60%; Asian = 24.00%; White, non-Hispanic = 14.07%; Sharpless, 2019).

Further evidence of workforce diversity constriction emerged in Puerto Rico when attempting to translate the EPPP into Spanish. Puerto Rico became a member of ASPPB on the condition that a bilingual Spanish EPPP (S-EPPP) would be made available to applicants in that jurisdiction (Law 281-2012). ASPPB translated two forms of the EPPP into Spanish and began to offer them in Puerto Rico (Hilson A., 2016). However, as described in Law 183-2015, preparation of those forms was rushed, did not engage stakeholders, and did not include a sufficient validation process. After the launch, the failure rate on the S-EPPP was so high that it resulted in a workforce crisis across the jurisdiction and the earlier law was revoked (via Law 183-2015). ASPPB subsequently discontinued the S-EPPP entirely (ASPPB, 2016).

Smaller, but still discernible, differences have also been observed based on binary gender identification with men failing more often than women (18.85% versus 15.82%). We did not locate any data regarding disability status and EPPP scores, suggesting an area of needed research. However, we did find evidence that pass rates on the United States Medical Licensing Examination (USMLE), which serves a similar professional gatekeeping purpose, were lower for those with disabilities (Meeks & Herzer, 2015). Similar inquiries in other healthcare professions, such as psychology, may be informative in understanding the possible role of credentialing examinations in the underrepresentation of individuals with disabilities in the workforce.

**Consequences for doctoral training.** Based on the observation that candidates pass at a higher rate when the current EPPP is taken closer to the completion of doctoral coursework (Schaffer et al., 2012), ASPPB infers that it would be beneficial for students to take the current knowledge exam (Part 1) of the Enhanced EPPP before completing dissertation and internship requirements. Hence, when Part 2 is introduced, timing of Part 1 is expected to shift. There are multiple conceptual and pragmatic complications for doctoral training that are associated with this inference but, for efficiency, we will draw attention to two conceptual problems that may be less obvious.

First, assessment of the knowledge base in psychological science and clinical application training is regulated by doctoral program accreditation (e.g., APA, 2018; PCSAS, 2011), with the understanding that competency-based evaluations are the responsibility of doctoral programs, not ASPPB. While ASPPB can, and should, participate as a stakeholder in doctoral accreditation regulation processes, ASPPB assessment of this same body of knowledge is unjustified. Observation that exam pass rate likelihood is tied to proximity of doctoral coursework (Shaffer et al., 2012), coupled with findings that pass rates are higher among students graduating from accredited programs (McGaha & Minder 1993; Ross, Holzman, Handal, & Gilner, 1993; Schaffer et al., 2012; Templer & Tomeo, 1998; Templer & Tomeo, 2000; Yu et al., 1997), suggests doctoral programs and their accrediting bodies are performing their duties well. Indeed, the strong correspondence between timing of doctoral coursework completion and success on the Part 1 knowledge exam suggests unnecessary evaluative redundancy. In addition, trainees will be subjected to preparing for Part 1 of a licensure exam that overlaps with required preparation and testing for high stakes program requirements that include, but are not limited to, qualifying and comprehensive examinations as well as the dissertation proposal defense.

Second, placement of Part 1 testing within doctoral training may significantly undermine the quality of doctoral training itself. Exam preparation time seems highly likely to draw time away from research, teaching, and/or practicum training that is necessary for development of these skills and abilities. To preserve time for those experiences, programs may feel pressed (by students and/or faculty) to teach to the test. There is an additional risk that some internship sites could screen applicants by stratifying them according to their Part 1 exam scores. As described earlier, it has already been demonstrated that Part 1 scores are associated with unintended racial/ethnic referents (APA Commission on Ethnic Minority Recruitment, Retention, and Training in Psychology Task Force, 2008; Bowman & Ameen, 2018) as well as gender (Sharpless, 2019). As such, internship placements could quickly begin to inadvertently stratify by demographic variables if applicants are screened by test score.

**Inconsistencies in implementation across states.** If both parts of the Enhanced EPPP are retained, the timing of the two exams will be determined by each state licensing board, with exam readiness requirements likely to vary across jurisdictions (as they do now). Such variability may inadvertently contribute to inequities by enabling trainees from jurisdictions with earlier exam completions to be more rapidly available to compete for employment. (We note that such variability is not found in the medical licensure process, which itself is conducted by licensing boards in each state.) Differences in licensing guidelines would also likely further hinder professional mobility. Sample scenarios for consideration: Will it be necessary to retake the EPPP Part 1 if a candidate is seeking licensure in a state that requires completion of both sections at the postdoctoral level? Will jurisdictions with less restrictive limits on access to the EPPP Part 1 receive a disproportionate increase in the number of potential licensees who indicate an intent to practice in that state? Will federal employees disproportionately seek and maintain licenses in

states where they do not reside to avoid local regulations? Each of these requires careful consideration. At a minimum, the current implementation plans for the Enhanced EPPP appear likely to amplify existing problems related to jurisdictional inconsistency rather than resolve them. Likewise, there is no evidence to date that a majority of jurisdictions are satisfied with the validation process and willing to incorporate the new test into their licensing procedures.

**Personal financial burden.** Doctoral training in health service psychology typically results in significant debt burden by the time of licensure eligibility (Doran, Kraha, Marks, Ameen, El-Ghoroury, 2016). Fees for test registration and exam preparation materials are scheduled to increase substantially with the advent of the Enhanced EPPP's two-part examination model. Although ASPPB indicates efforts to contain fee increases associated with the Enhanced EPPP, it seems inevitable that most exam costs will double relative to current levels. In addition to direct expenses, it is likely that licensure applicants will incur indirect costs as well, such as lost productivity and income and greater debt as employment is delayed while trainees prepare for the longer, two-part exam. Of import, available evidence suggests that student diversity status, debt load, and likelihood of passing the current EPPP intersect (Bowman & Ameen, 2018; Doran et al., 2016; Sayette, Norcross, & Dimoff, 2011), raising concerns that the Enhanced EPPP will further limit access to the field for trainees from diverse racial, ethnic, or socioeconomic backgrounds.

**Possible legal challenges.** “According to ASPPB, the EPPP Part 2 is an assessment of skills or application of knowledge in a manner that is reliable, valid and defensible” (H. Broadus, personal communication, 2/5/2019 from the Deputy Director of the Arizona Board of Psychologist Examiners). Since ASPPB has apparently been understood as asserting defensibility with a member jurisdiction, consideration of legal standards for high stakes

assessment is worth brief consideration. According to the Supreme Court (*Daubert v. Merrell Dow Pharmaceuticals*, 1993) there are four specific qualities applied to tests being introduced in a legal context that determine the admissibility of scientific evidence. All four of the following conditions must be met: testability, peer-review and publication, error rate, and general acceptance. Based on what is known at this time, it is not clear whether the Enhanced EPPP meets any of the *Daubert* requirements.

In written and verbal statements, ASPPB has used a variety of terms to describe their hypothesis regarding what the Enhanced EPPP, particularly Part 2, is designed to evaluate. Developed under the auspices of competency-based evaluation, Part 2 is commonly described as “a test of skills” or “an integrated test of knowledge and skills” while “a test of applied knowledge” has also been acknowledged (J. Horn, personal communication, January 18, 2019, 2019 CUDCP Mid-Winter Meeting). Such a multiplicity of descriptions suggests a lack of clear theoretical grounding or falsifiability and, consequently, potential failure to meet the testability requirement. ASPPB’s exclusive reliance on unpublished research that has not been subjected to scholarly peer review falls short of meeting the second requirement of peer review and publication (Reisberg, Simons, & Fournier, 2016). The lack of external validity evidence precludes identification of an error rate (Faigman, 2013; Gatowski et al., 2001; Meixner & Diamond, 2014), which violates the third requirement.

The fourth and final requirement, general acceptance, refers specifically to acceptance by the relevant scientific and professional communities. A letter to ASPPB, dated October 5, 2018, indicates that Part 2 has not attained general acceptance by relevant scientific communities. The letter sent on behalf of 10 councils and professional organizations<sup>5</sup> that collectively represent

---

<sup>5</sup> In alphabetical order, the ten co-signing organizations/councils to that letter were as follows: Academy of Psychological Clinical Science; Association of Counseling Center Training Agencies, Association of Postdoctoral



approximately 800 doctoral training programs or clinics in health service psychology, spanning all licensure jurisdictions, details multiple concerns associated with the construction and implementation of Part 2. To summarize, the letter states that the councils and organizations, “remain deeply concerned that the measure development process is inadequate and the planned launch of the Enhanced EPPP is premature. We strongly urge ASPPB to (1) address the many stakeholder concerns regarding development of a high-quality, valid, and accessible exam, (2) not hesitate to slow and alter exam development to fully resolve these challenges and concerns, and (3) more fully involve stakeholders, including state associations and training councils, in constructive dialogue and data collection that resolves these issues.” In fact, some jurisdictions have chosen not to adopt the EPPP Part 2; including, most recently, New York (New York State Education Department, Office of the Professions, Board for Psychology, personal communication to doctoral programs, September 4, 2019).

### **Altering the Trajectory**

It is ultimately the responsibility of jurisdictions to determine how and when to implement any changes to licensure examinations. Jurisdictions are not only within their scope of authority to make all implementation decisions, it is their duty to make reasoned, evidence-based decisions that protect the public’s interests. In forming expectations around what constitutes sufficient validation prior to implementation, jurisdictions may find it useful to consider how other health care professions have approached the need to modernize licensure examinations in their disciplines and mitigate shortcomings that potentially jeopardize the best interests of consumers in their jurisdictions.

---

Programs in Clinical Neuropsychology, Association of Psychology Training Clinics, Clinical Child and Pediatric Psychology Training Council, Consortium of Combined-Integrated Doctoral Programs in Psychology; Council of Counseling Psychology Training Programs; Council of Directors of School Psychology Programs; Council of Graduate Departments of Psychology; and Council of University Directors of Clinical Psychology.

### **Lessons from the Assessment of Licensure Readiness in Other Health Service Disciplines**

The National Board of Medical Examiners (NBME; 2019), responsible for competency assessment in medicine, has taken a notably different approach to the evaluation of licensure readiness than that of ASPPB. Here, we focus on four particularly critical points of contrast that we believe have worked well for medicine and could provide guidance for psychology: engagement of stakeholders; peer review; breadth of construct validity evidence; and a broader use of validation processes.

**Engagement of stakeholders.** Stakeholder engagement permeates licensure evaluation much more fully in medicine than in psychology. Consider, for example, that ASPPB's membership consists of representatives from state and provincial licensing boards and that the organization serves as the sole owner and operator of the licensing exam. Input from stakeholders outside of jurisdictional regulatory bodies – on ASPPB's functioning, licensure evaluation processes, and the licensing exam itself – is typically in the form of responses to occasional and fairly specific requests for comments solicited by ASPPB, or informal feedback provided through ASPPB's liaison relationships with other professional organizations. In contrast, NBME includes representation from national experts who contribute to the design of its examinations, at-large members who include members of the public, and representatives from multiple stakeholder organizations.<sup>6</sup> This representation is remarkably broader than ASPPB's,

---

<sup>6</sup> NBME includes representatives from the Association of American Medical Colleges, the American Board of Medical Specialties, the American Medical Association (AMA), the Resident and Fellows Section of the AMA, the American Medical Student Association, the Council of Medical Specialty Societies, the Educational Commission for Foreign Medical Graduates, the Federation of State Medical Boards (FSMB), Student National Medical Association, the uniformed services, and the Department of Veterans Affairs. Similarly, the Joint Commission on National Dental Examinations (JCND E) is the agency responsible for the development and administration of the National Board Dental Examinations (NBDE). Their 15-member Commission includes representatives from dental schools, dental practice, state dental examining boards, dental hygiene, dental students, and the public (see description here: <https://www.ada.org/en/jcnde/about-us/jcnde-members-and-appointing-organizations>).

with a formal structure that ensures broad representation and inclusion of stakeholders in the profession.

Competency assessment for medical licensure is similarly managed by a collaborative stakeholder group. The United States Medical Licensing Examination (USMLE; 2019) is governed by members from the NBME, the Federation of State Medical Boards, the Educational Commission for Foreign Medical Graduates, and the public. This governing committee is responsible for overseeing all significant policies and procedures, including maintaining the quality of the assessment process, identifying potential conflicts of interest that could interfere with widespread acceptance of the test, determining the overall direction of the program, and identifying and approving procedures for scoring and determining the pass/fail standard. Medicine's reliance on a structured checks-and-balances approach, with three partnering bodies that work in tandem, formalizes broad and ongoing stakeholder involvement in the regulation of their profession. A similar governing structure for licensure examination might be very beneficial to psychology.

The USMLE test content is determined by a set of test committees appointed by the overall committee. Broad stakeholder involvement is evident here as well; the test committees include biomedical scientists, medical educators, and clinicians, and virtually all accredited medical schools in the United States have been represented on one or more USMLE test committees. USMLE test committee members represent a "national faculty of medicine" drawn from medical schools, state medical boards, and clinical practice settings across the United States ("Who is USMLE?", 2019).

**Peer review.** Peer review has been described as the bedrock of quality control in the field of psychological science (e.g., Reisberg, et al, 2016). Peer-review allows for rigorous evaluation

of validity and statistical reliability (Faigman, 2013; Gatowski et al., 2001). However, to our knowledge, there is no involvement of independent investigators to promote quality science regarding the EPPP Part 2. Further, at the time of this writing, analyses and findings associated with the EPPP Part 2 have not been subjected to peer review.

In contrast, the NBME has an explicit commitment to make test data available to researchers to perform independent examinations of test validity and reliability and to conduct research on medical assessment and competency using the dataset. The NBME Data Sharing and Research Collaboration Program provides test score data and related information to appropriately vetted external investigators to pursue topics of their own interest that will also benefit the health professions education community or measurement community by expanding knowledge and improving practice. Through this program, NBME promotes research and evaluation in assessment by building relationships with academic and applied researchers and the organizations with which they are affiliated. It is expected that completed research will be submitted for publication in peer-reviewed journals. A similar commitment to independent examination of the EPPP that can withstand the rigors of peer review would almost certainly yield a wealth of invaluable information that could help our profession shape a strong licensure evaluation process.

**Breadth of construct validity evidence.** The IoM Committee on the Health Professions Education (2003) recommends that “individual disciplines should benefit from each other’s knowledge and experience in creating and implementing more valid and reliable assessment[s] of competency for initial licensure and beyond.” In keeping with this emphasis on more valid and reliable assessments, the NBME systematically develops plans to enhance assessment of competency with the full cooperation of relevant stakeholders. Current projects include

investigations related to the assessment of new constructs and competencies, simulations and performance testing, test score scaling and equating, score reporting and feedback, validity of test score use, group differences, and general psychometrics (NBME, 2019). Notably, all of those projects are equally appropriate to health service psychology and could be used to expand and enrich ASPPB's current emphasis on content validity. Although instituting such a broad scope of activities is labor intensive and expensive, many of these goals can be readily achieved within psychological science given the research interests and expertise represented in our ranks. Examples involving two domains, score correspondence and incremental validity, are discussed further below.

**Use of validation processes.** Consistent with other health profession accreditation groups, the USMLE incorporates four working principles to assessment for medical licensure: (a) continually determining the acceptability of the program to stakeholders, (b) encouraging stakeholders to participate actively in evaluating and improving the test, (c) monitoring and evaluating the correspondence between performance on the test and relevant external measures of competency in medical practice, and (d) using the findings from that evaluation process to revise and improve the test itself over time (Epstein & Hundert, 2002). This overall process reflects the USMLE evolution in approach over the past 20 years -- initiating strategic steps to move away from exclusive reliance on job task analysis and content validity in the test development process, and moving toward an iterative multi-step sequence of content development and evaluation in cooperation with multiple stakeholders. In this way, the USMLE has engaged in a process of validation that includes careful analysis of the function of the test for its intended purpose. (For discussions of how other disciplines have taken on similar challenges successfully, see Gadbury-Amyot, McCracken, Woldt, & Brennen, 2014; Rose & Regan-

Kubinski, 2010). It is probable that Part 2 of the EPPP could similarly benefit from iterative validation studies.

### **Specific Recommendations for Jurisdictions Licensing Health Service Psychologists**

Thus far, we have made several specific recommendations concerning additional research that should be conducted, and we have advocated for much greater infusion of stakeholder feedback into the exam creation and governance processes. For the more immediate future, we offer several recommendations to jurisdictions regarding the Enhanced EPPP. We strongly encourage jurisdictions to look for three key indicators of readiness prior to making any implementation decisions regarding the Enhanced EPPP: (a) greater depth and breadth in psychometric inquiries of the examination; (b) peer review of each study cited as contributing to implementation recommendations; and (c) broader participation of, and acceptance by, relevant stakeholders, including appropriate scientific communities. The first and second indicators are self-evident via review of citations (e.g., verifying psychometric inquiries are published in peer reviewed journals). The third indicator could be evaluated by seeking explicit, formal input from the broad stakeholder base (e.g., training councils; other professional organizations that are independent of ASPPB). With respect to psychometric inquiries, we provide a sampling of possible studies in two different areas to illustrate the breadth and depth of scope that is feasible via independent investigations (as is done in medicine) and/or in cooperation with ASPPB. By no means is this an exhaustive listing of feasible or necessary work but rather, these exemplars are offered to encourage greater discussion and involvement by stakeholders and independent investigators.

**Score correspondence.** Prior to adoption, jurisdictions will need information on whether scores from the Enhanced EPPP correspond to intended referents (e.g., competency) and not

unintended referents (e.g., race; socioeconomic status; gender identification; disability status; national origin). Multiple yardsticks for the assessment of score correspondence with appropriate referents could be considered. For example, a known groups design could be used to examine whether there are differential pass rates among those holding unrelated doctoral degrees (e.g., engineering), related sub-doctoral degrees (e.g., master's degree in health service psychology), and doctoral degrees in health service psychology. Previous research underscores the likelihood of discernible exam performance differences (e.g., master's versus doctoral levels; DeMers, 2009) and may provide important information to jurisdictions in determining the appropriate scope of practice for doctoral level health service psychology, particularly in light of forthcoming accreditation of master's programs in health service psychology (Callahan, 2019).

For the original EPPP, ASPPB's examination committee, the test vendor, and individual jurisdictions worked with independent investigators to facilitate validation inquiries involving known groups designs. Included in these designs were undergraduate students (Shrader, 1979; Wertheimer, 1972; 1974), master's level individuals seeking licensure (Hays, 1970; Loveless, 1979; Terris, 1973), doctoral level individuals according to degree type (e.g., PhD, PsyD, and/or EdD) (Hays & Mullins, 1978; Hays & Schreiner, 1977), and individuals failing the exam at least once and as many as four times (Brown, 1979). We are unable to verify all of the primary sources for those inquiries since some rely on communications from investigators that were sent directly to ASPPB. However, Shrader (1980) reportedly had access to all of these reports and concluded that known groups fell into a hierarchy of mean exam scores as follows: Ph.D. Psychology, Psy.D. Psychology, Ed.D. Psychology, Master's Degree Psychology, Graduate Students in Psychology; Undergraduate Honors Students in Psychology; Other Psychology Undergraduates. Shortly thereafter, Hoffman (1980) reported similar findings in which mean

EPPP scores fell into a hierarchy based on the type of degree institution: Major University; Small or Unknown College; Professional School. Such differences were evidenced not only at the total score level, but also when examining subtest scores (Templer & Tomeo, 1998; Templer & Tomeo, 2000). A known groups design would also allow for tests of independence between Part 2 examination scores and unintended referents (e.g., diversity variables).

As noted earlier, the field of medicine expects licensure scores to correspond with competent skills performance in other settings. Of importance for research of that type, work in our discipline has already gone into creating psychometric scales to measure demonstrable competency while accounting for the phenomenon of rater biases (e.g., Price, Callahan, & Cox, 2017). Such scales, in concert with standardized simulated patient scenarios (e.g., Cramer, Johnson, McLaughlin, Rausch, & Conroy, 2013; Sheen, McGillivray, Gurtman, Boyd, 2015), could be used to investigate score correspondence. To be clear, we are not suggesting all examinees participate in simulated patient scenarios, which is presently beyond the financial and logistic resources of our profession and licensure applicants. Rather, we are suggesting focused research studies that could, with adequate statistical power, contribute important information to the validation process and inform decisions concerning implementation.

**Incremental validity.** Dismantling designs and research on incremental validity would allow jurisdictions to parse the Enhanced EPPP into the discrete contributions of Part 1 and Part 2. Although ASPPB does not support such inquiries (J. Hunsley, personal communication, December 7, 2017) and holds an *a priori* belief that all components of the Enhanced EPPP are essential, there is no empirical evidence at this time to support such an assertion. In fact, adopting such a perspective appears counter to the implied framing of the Enhanced EPPP Part 2.



ASPPB has publicly stated<sup>7</sup> that ~95% of those who pass Part 1 are likely to also pass Part 2. This suggests that the Enhanced EPPP is not likely to demonstrate any significant incremental validity, which underscores our concerns of excessive and unnecessary redundancy. Based on what ASPPB has stated, we hypothesize that studies of incremental validity would show no incremental benefit with the use of both exams and therefore recommend that only the superior of the two exams be implemented. Unfortunately, we cannot be sure at this time which of the two examinations (Part 1 or Part 2) is superior.

### **Conclusions**

While ASPPB has engaged in an intensive and ambitious process to develop test items for the Enhanced EPPP Part 2 that reflect core job tasks, they have been less engaged in the fuller process of validation, and have not communicated an openness to greater engagement from relevant stakeholders, involvement of psychological experts, and peer review. As a result, the overall Enhanced EPPP development process is less sophisticated and comprehensive than similar undertakings in other professions, which could have untoward consequences for health service psychology and the public, and leaves open the door for legal challenges regarding the Enhanced EPPP.

As noted above, the recent history of the USMLE highlights important trends in assessment for licensing in health care provision (e.g., Elstein, 1993; Epstein & Hundert, 2002). We find it particularly salient that other disciplines have relied heavily on strict and comprehensive application of psychological principles and psychometric theory in tackling such challenging questions as the development of optimal and efficient methods for assessing clinical reasoning (Longo, Orcutt, James, Kane, & Coleman, 2018; Rencic, Durning, Holmboe, &

---

<sup>7</sup> Addressing the full body of attendees at the January 2019 Mid-Winter Meeting of the Council of University Directors of Clinical Psychology.

Gruppen, 2016) and for ensuring the predictive validity of competency assessment for health care practice (Melnick & Clauser, 2005). It is encouraging and reassuring to see how psychological knowledge and assessment practices have led to substantial improvements in assessment for health care licensure and practice. Both pragmatic and legal concerns have driven those involved with competency evaluation for medical licensure and practice to invest significant resources, to collaborate extensively and, in the process, to focus more intensively on validity and reliability (Govaerts & van der Vleuten, 2013; Norcini, et al., 2011).

ASPPB's process to date has not reflected the same level of investment in validation studies or in strategic inclusion of stakeholders and researchers with relevant expertise. We have every reason to believe that relevant stakeholders would be willing to engage productively in this process as partners with the common goal of ensuring public health and well-being. As a discipline, we must compel ourselves to apply rigorous methods and standards to putative developments in the evaluation of those seeking licensure in health service psychology. To hold ourselves less accountable to our own standards and ethics than other disciplines – who cite our discipline's standards as foundational to their work – is disappointing at best and self-defeating at worst. Our field has the knowledge and skills to produce a well-validated and appropriate licensure assessment process, as well as stakeholders who are clearly committed to participating in the validation process. It is imperative that we use these resources fully to protect the discipline and the public.

To be clear, we do not recommend abandoning Part 2 development. As noted earlier, independent, peer-reviewed studies repeatedly report significant limitations associated with the existing EPPP, particularly with respect to diversity constriction (e.g., Bowman & Ameen, 2018; Sharpless, 2019; Sharpless & Barber, 2009). Thus, based on the rigor of the work force analysis

and attention to construct validity described thus far, we strongly suspect Part 2 carries potential to emerge as psychometrically superior to the current EPPP. Rather than abandon development of Part 2, we encourage more thorough investment in its validation in concert with planned obsolescence for the current examination (i.e., Part 1). We strongly urge ASPPB and jurisdictions to partner with additional stakeholders and qualified investigators to facilitate the needed validation studies, including examination of a) structural and measurement invariance and b) incremental validity to re-evaluate the necessity of both exams.

Until these goals are reached, we urge ASPPB and licensure jurisdictions to hold off on implementation of the Enhanced EPPP. Potential costs of implementing a test prematurely include costs to trainees in time and burden, costs to the discipline of implementing an additional gate that may disproportionately influence more vulnerable trainees, and the potential for state boards to face legal challenges regarding their licensure standards.

## References

- Agency for Healthcare Research and Quality (2017). *2016 National Healthcare Quality and Disparities Report. AHRQ Pub. No. 17-0001*. Rockville, MD: Author.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: Author.
- American Psychological Association (2008). *Guidelines and principles for accreditation of programs in professional psychology*. Washington, D.C.: Author.
- American Psychological Association (2018). *Standards of Accreditation for Health Service Psychology*. Washington D.C.: American Psychological Association.
- American Psychological Association, Office of Ethnic Minority Affairs. (2008). *A portrait of success and challenge—Progress report: 1997–2005*. Washington, DC: Author.  
Retrieved from <https://www.apa.org/pi/oema/resources/success-challenge.pdf>
- Association of State and Provincial Psychology Boards. (2017). *The EPPP part 2: The assessment of skills needed for the independent practice of psychology*. Retrieved from: <https://www.asppb.net/page/EPPPPart2>
- Association of State and Provincial Psychology Boards. (2016, November 10). *Spanish/English EPPP*. Retrieved from: <https://www.asppb.net/news/317115/SpanishEnglish-EPPP.htm>
- Bowman, N. & Ameen, E. (2018). *Exploring differences in pass rates on the Examination for Professional Practice in Psychology*. Retrieved from <https://www.apa.org/pi/oema/resources/communique/2018/06/pass-rates>
- Brown, J. M. (1979). Analysis of data from candidates taking the EPPP. Personal communication with Shrader (1980).

- Callahan, J. L. (2019). Master's level accreditation in health services psychology: A primer to the special section with commentary. *Training and Education in Professional Psychology, 13*, 73-83. doi: 10.1037/tep0000248
- Callahan, J. L., Smotherman, J., Dziurzynski, K., Love, P. K., Davis, E. C., Flores Niemann, Y., & Ruggero, C. J. (2018). Diversity in the professional psychology training-to-workforce pipeline: Results from doctoral psychology student population data. *Training and Education in Professional Psychology, 12*, 273-285. doi: 10.1037/tep0000203
- Callahan, J. L. & Watkins, Jr., C. E. (2018a). Evidence-based training: The time has come. *Training and Education in Professional Psychology, 12*, 211-218. doi: 10.1037/tep0000204
- Callahan, J. L. & Watkins, Jr., C. E. (2018b). The science of training III: Supervision, competency, and internship training. *Training and Education in Professional Psychology, 12*, 245-261. doi: 10.1037/tep0000208
- Cizek, G. J. (2012). Defining and distinguishing validity: Interpretations of score meaning and justifications of test use. *Psychological Methods, 17*, 31-43. doi: 10.1037/a0026975
- Cizek, G. J. (2016). Validating test score meaning and defending test score use: Different aims, different methods. *Assessment in Education: Principles, Policy & Practice, 23*, 212-225. doi: 10.1080/0969594X.2015.1063479
- Cramer, R. J., Johnson, S. M., McLaughlin, J., Rausch, E. M. & Conroy, M. A. (2013). Suicide risk assessment training for psychology doctoral programs: Core competencies and a framework for training. *Training and Education in Professional Psychology, 7*, 1-11. doi: 10.1037/a0031836
- Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U. S., 113 S. Ct. 2786 (1993).

- DeMers, S. T. (2009). Understanding the purpose, strengths, and limitations of the EPPP: A response to Sharpless and Barber. *Professional Psychology: Research and Practice, 40*, 348-353. doi: 10.1037/a0015750
- DiLillo, D. & Tremblay, G. C. (2009). How should the effectiveness of the EPPP be judged? *Professional Psychology: Research and Practice, 40*, 345-347. doi: 10.1037/a0015734
- Doran, J. M., Kraha, A., Marks, L. R., Ameen, E. J., & El-Ghoroury, N. H. (2016). Graduate debt in psychology: A quantitative analysis. *Training and Education in Professional Psychology, 10*, 3-13. doi: 10.1037/tep0000112
- Elstein, A. S. (1993). Beyond multiple-choice questions and essays: The need for a new way to assess clinical competence. *Academic Medicine, 68*, 244-249. doi: 10.1097/00001888-199304000-00002
- Epstein, R.M., & Hundert, E.M. (2002). Defining and assessing professional competence. *Journal of the American Medical Association, 287*, 226-235. doi:10.1001/jama.287.2.226
- Erikson Cornish, J. A. & Smith, R. D. (2009). Reflections on the EPPP: A commentary on Sharpless and Barber. *Professional Psychology: Research and Practice, 40*, 341-344. doi: 10.1037/a0015412
- Faigman, D. L. (2013). The *Daubert* revolution and the birth of modernity. Managing scientific evidence in the age of science. *University of California at Davis Law Review, 46*, 100-138.
- Gadbury-Amyot, C. C., McCracken, M. S., Woldt, J. L., & Brennan, R. L. (2014). Validity and reliability of portfolio assessment of student competence in two dental school populations: a four-year study. *Journal of Dental Education, 78*, 657-667.

- Gatowski, S. I., Dobbin, A. D., Richardson, J. T., Ginsburg, G. P., Merlino, M. L., & Dahir, V. (2001). Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-*Daubert* world. *Law and Human Behavior, 25*, 433-458. doi: 10.1023/A:1012899030937
- Gervais, J. (2016). The operational definition of competency-based education. *The Journal of Competency-Based Education, 1*, 98-106. doi: 10.1002/cbe2.1011
- Govaerts, M., & van der Vleuten, C. P. (2013). Validity in work-based assessment: Expanding our horizons. *Medical Education, 47*, 1164-1174. doi: 10.1111/medu.12289
- Hays, J. R. (1970). Report to the Executive Committee, American Association of State Psychology Boards. Reported in Shrader (1980).
- Hays, J. R. & Mullins, D. (1978). Background characteristics of psychologists taking the examination for professional practice in psychology. *Texas Psychologist, 30*, 1.
- Hays, J. R., & Schreiner, D. (1977). Comparison of degree received and performance on licensing examination. *Psychological Reports, 40*, 42. doi: 10.2466/pr0.1977.40.1.42
- Hilson, A. (2016, Spring). Executive message: The ASPPB examination program. *ASPPB Central*. Retrieved from: [https://cdn.ymaws.com/www.asppb.net/resource/resmgr/Guidelines/ASPPB\\_Central\\_Spring\\_Edition.pdf](https://cdn.ymaws.com/www.asppb.net/resource/resmgr/Guidelines/ASPPB_Central_Spring_Edition.pdf)
- Hoffman, P. J. (1980). Validation of the EPPP: The California pilot study. *Professional Practice of Psychology, 1*, 31-44.
- Hubley, A. M., & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research, 103*, 219. doi: 10.1007/s11205-011-9843-4

- Hughes, D. J. (2018). Psychometric validity: Establishing the accuracy and appropriateness of psychometric measures. *The Wiley handbook of psychometric testing: A multidisciplinary approach to survey, scale and test development*. Chichester, UK: Wiley.
- Ibaraki, A. Y., & Hall, G. N. (2014). The components of cultural match in psychotherapy. *Journal of Social and Clinical Psychology, 33*, 936-953. doi: 10.1521/jscp.2014.33.10.936
- Institute of Medicine, Board on Health Care Services, Committee on the Health Professions Education Summit (2003). Health professions oversight processes: What they do and do not do, and what they could do. *Health professions education: A bridge to quality* (A.C. Greiner & E. Knebel, Eds.), pp. 97-120. Washington (DC): National Academies Press (US). Available from: <http://www.nap.edu/catalog/10681.html>
- Kane, M. T. (2016). Validity as the evaluation of the claims based on test scores. *Assessment in Education: Principles, Policy & Practice, 23*, 309-311. doi: 10.1080/0969594X.2016.1156645
- Lees-Haley, P. R. (1996). Alice in Validityland, or the dangerous consequences of consequential validity. *American Psychologist, 51*, 981-983. doi: 10.1037/0003-066X.51.9.981
- Longo, P. J., Orcutt, V. L., James, K., Kane, J., & Coleman, V. (2018). Clinical reasoning and knowledge organization: Bridging the gap between medical education and neurocognitive science. *The Journal of Physician Assistant Education, 29*, 230-235. doi: 10.1097/JPA.0000000000000224
- Loveless, H. E. (1979). Personal communication with Shrader (1980).



- McGaha, S. & Minder, C. (1993). Factors influencing performance on the Examination for Professional Practice in Psychology. *Professional Psychology: Research and Practice*, 24, 107-109. doi: 10.1037/0735-7028.24.1.107
- Meeks, L. M. & Herzer, K. R. (2015). Accommodated MCAT time and performance standards. *JAMA: Journal of the American Medical Association*, 314, 1517-1518. doi: 10.1001/jama.2015.10741
- Melnick, D. E., & Clauser, B. E. (2005). Computer-based testing for professional licensing and certification of health professionals. In D. Bartram & R. Hambleton (Eds.), *Computer-based testing and the Internet: Issues and advances* (pp. 163-197). New York: Wiley.
- Meixner, J. B. & Diamond, S. S. (2014). The hidden *Daubert* factor: How judges use error rates in assessing scientific evidence. *Wisconsin Law Review*, 2014, 1063-1133.
- National Board of Medical Examiners (2019). <https://www.nbme.org/about/index.html>
- Newton, P. E. (2012). Clarifying the consensus definition of validity. *Measurement: Interdisciplinary Research & Perspective*, 10, 1-29. doi: 10.1080/15366367.2012.669666
- Norcini, J. J., Lipner, R. S., & Grosso, L. J. (2013). Assessment in the context of licensure and certification. *Teaching and Learning in Medicine*, 25(S1), S62-S67. doi: 10.1080/10401334.2013.842909
- O'Connell, Jr., W. R., & Moomaw, E. W. (1975). *A CBC Primer. Report of a Conference: Competency-Based Curricula in General Undergraduate Programs*. Retrieved from <http://eric.ed.gov/?id=ED104297>
- Owen, J., Tao, K. W., Imel, Z. E., Wampold, B. E., & Rodolfa, E. (2014). Addressing racial and ethnic microaggressions in therapy. *Professional Psychology: Research and Practice*, 45(4), 283-290. doi:10.1037/a0037420

- Price, S. D., Callahan, J. L., Cox, R. J. (2017). Psychometric investigation of competency benchmarks. *Training and Education in Professional Psychology, 11*, 128-139. doi: 10.1037/tep0000133
- Psychological Clinical Science Accreditation System (2011). Accreditation review Criteria. <https://www.pcsas.org/accreditation/review-criteria/>
- Act No. 193, Puerto Rico § 193-2015 (2015). Retrieved from: <http://www.oslpr.org/2013-2016/leyes/pdf/ley-193-20-Nov-2015.pdf>
- Act No. 281, Puerto Rico § 281-2012 (2012). Retrieved from: <http://www.oslpr.org/2009-2012/leyes/pdf/ley-281-29-Sep-2012.pdf>
- Reisberg, D., Simons, D. J., & Fournier, L. R. (2016). Introduction to the forum on when and whether psychological research is ready for use in the justice system. *Journal of Applied Research in Memory and Cognition, 5*, 233-235. doi: 10.1016/j.jarmac.2016.07.009
- Rencic J., Durning, S. J., Holmboe, E., & Gruppen, L. D. (2016). Understanding the assessment of clinical reasoning. In P. Wimmers & M. Mentkowski (Eds.), *Assessing competence in professional performance across disciplines and professions. Vol. 13, Innovation and change in professional education* (pp. 209-235). New York: Springer.
- Rose, L., & Regan-Kubinski, M. J. (2010). Update on advanced practice registered nurse regulation: Licensure, accreditation, certification, and education. *Archives of Psychiatric Nursing, 24*, 440-451. doi: 10.1016/j.apnu.2010.08.003
- Rosen, G. A., Reaves, R. P., & Hill, D. S. (1989). Reliability and validity of psychology licensing exams: Multiple roles and redundant steps in development and scoring, *Professional Psychology: Research and Practice, 20*, 272-274. doi: 10.1037/0735-7028.20.4.272

- Ross, M. J., Holzman, L. A., Handal, P. J., & Gilner, F. H. (1991). Performance on the Examination for Professional Psychology as a function of specialty, degree, administrative housing, and accreditation status. *Professional Psychology, 22*, 347-350. doi: 10.1037/0735-7028.22.5.347
- Ryan, A. M., & Chan, D. (1999). Perceptions of the EPPP: How do licensure candidates view the process? *Professional Psychology: Research and Practice, 30*, 519-530. doi: 10.1037/0735-7028.30.5.519
- Sayette, M. A., Norcross, J. C., & Dimoff, J. D. (2011). The heterogeneity of clinical psychology Ph. D. programs and the distinctiveness of APCS programs. *Clinical Psychology: Science and Practice, 18*, 4-11. doi: 10.1111/j.1468-2850.2010.01227.x
- Schaffer, J., Rodolfa, E., Owen, J., Lipkins, R., Webb, C. & Horn, J. (2012). The Examination for Professional Practice in Psychology: New data—practical implications. *Training and Education in Professional Psychology, 6*, 1-7. doi: 10.1037/a0026823
- Sharpless, B. A. (2019). Are demographic variables associated with performance on the Examination for Professional Practice in Psychology (EPPP)?. *Journal of Psychology, 153*, 161-172. doi: 10.180/00223980.2018.1504739
- Sharpless, B. A., & Barber, J. P. (2009). The Examination for Professional Practice in Psychology (EPPP) in the era of evidence-based practice. *Professional Psychology: Research and Practice, 40*, 333. doi: 10.1037/a0013983
- Sheen, J., McGillivray, J., Gurtman, C., & Boyd, L. (2015). Assessing the clinical competence of psychology students through Objective Structured Clinical Examinations (OSCEs): Student and staff views. *Australian Psychologist, 50*, 51-59: doi: 10.1111/ap.12086

- Shepard, L. A. (2016). Evaluating test validity: Reprise and progress. *Assessment in Education: Principles, Policy & Practice*, 23, 268-280. doi: 10.1080/0969594X.2016.1141168
- Shrader, R. R. (1979). *Correlation of EPPP and GRE score*. Unpublished report presented to the Committee on Examination Procedures, American Association of State Psychology Boards. Reported in Shrader (1980).
- Shrader, R. R. (1980). Validation studies on the Examination for Professional Practice in Psychology. *Professional Practice of Psychology*, 1, 23-30.
- Sireci, S. G. (2016). On the validity of useless tests. *Assessment in Education: Principles, Policy & Practice*, 23(2), 226-235. doi: 10.1080/0969594X.2015.1072084
- Tao, K. W., Owen, J., Pace, B. T., & Imel, Z. E. (2015). A meta-analysis of multicultural competencies and psychotherapy process and outcome. *Journal of Counseling Psychology*, 62, 337-350. doi:10.1037/cou0000086
- Templer, D. I. & Tomeo, M. E. (1998). Examination for professional practice in psychology: Clerical errors corrected but inferences unchanged. *Psychological Sciences*, 9, 241-242. Doi: 10.1111/1467-9280.00048
- Templer, D. I. & Tomeo, M. E. (2000). Examination for professional practice in psychology subtest scores of professional and traditional clinical psychology program graduates. *The Journal of Psychology*, 124, 140-142. doi: 10.1080/00223980009600856
- Terris, L. D. (1973). The national licensing examination. *Professional Psychology*, 4, 386-391. doi: 10.1037/h0021436
- United States Medical Licensing Examination (2019). <https://www.usmle.org/about/>

- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-70. doi: 10.1177/109442810031002
- Wertheimer, M. (1972). *Scores of Psi Chi members on the EPPP*. Unpublished report presented to the Committee on Examination Procedures, American Association of State Psychology Boards. Reported in Shrader (1980).
- Wertheimer, M. (1974). *Scores of Psi Chi members on the EPPP*. Unpublished report presented to the Committee on Examination Procedures, American Association of State Psychology Boards. Reported in Shrader (1980).
- Who is USMLE? (2019, September 25). Retrieved from <https://www.usmle.org/about/>
- Wise, L. L. & Plake, B. S. (2016). Test design and development following the *Standards for Educational and Psychological Testing*. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of test development (2<sup>nd</sup> edition)* (pp. 19-39). New York: Routledge.
- Yu, L. M., Rinaldi, S. A., Templer, D. I., Colbert, L. A., Siscoe, K. & Van Patten, K. (1997). Score on the Examination for Professional Practice in Psychology as a function of attributes of clinical psychology graduate programs. *Psychological Science, 8*, 347-350. doi: 10.1111/j.1467-9280.1997.tb00423.x
- Zumbo, B. D., & Hubley, A. M. (2016). Bringing consequences and side effects of testing and assessment to the foreground. *Assessment in Education: Principles, Policy & Practice, 23*, 299-303. doi: 10.1080/0969594X.2016.1141169

**Table 1.** Types of Validity Evidence to Support Accuracy and Appropriateness, Frequency of Use, Application to the EPPP Part 2, and Planned Use by the ASPPB

Type of Evidence	Definition	% of Recent Test Development articles that used	Example relevant to EPPP Part 2	Described by ASPPB (2017) for the EPPP Part 2
<i>Accuracy Indices</i>				
Content	The degree to which the items of a measure comprehensively captures the target construct	55	Items match the types of competencies described in the job analysis survey	Yes
Structural	The degree to which the relationships among items reflect the theoretical framework (e.g., confirmatory factor analysis)	90	If the aim is to assess different aspects of competency, items for those different facets cohere into statistically distinct subscales	No <sup>8</sup>
Stability across groups	The degree to which the content and structural validity indicators are stable across groups	35	The degree to which items appear to relate to total scores in a similar manner for those from different SES, ethnic and racial backgrounds and gender	The goal is reported, but the details of which subgroups are being considered is not provided
<i>Appropriateness Indices</i>				
Convergent	The relationship of the new scale to other	75	The degree to which EPPP Part 2 scales relate to other	No

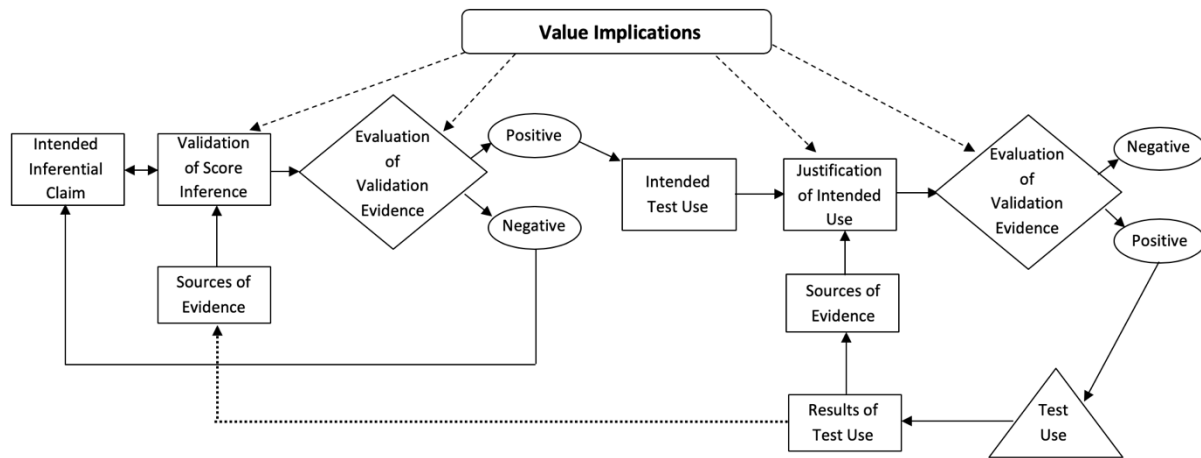
<sup>8</sup> ASPPB refers to an IRT approach consistently in their slideshows and accompanying narration, regardless of presenter. Unidimensionality is an assumption of IRT that would preclude examination of the 6-factor structure that ASPPB repeatedly assures is present in Part 2. It is entirely possible that the actual statistician employed by the test publisher working with ASPPB is using mIRT, multidimensional IRT, which would allow for such structural evidence. Unfortunately, ASPPB has appeared to be unwilling or unable to provide further information with respect to the IRT analyses.

	measures of the same construct		indices of competency, such as program evaluations of clinical competency	
Concurrent	Cross-sectional correlation of scores on the new test with criterion data collected at the same time	60	The degree to which EPPP Part 2 scales relate to completion of clinical psychology doctoral degree, to other indicators of skills in clinical psychology	No
Predictive	The ability of scores on the new test to longitudinally predict criterion scores	10	Over time, the degree to which EPPP Part 2 scores are found to relate to licensing board censure, client complaints, or dismissal from employment.	No. We concur with the ASPPB that in the absence of following those with failing scores, this would be difficult to attain.
Discriminant	The relationship of scores on the new test with scores on measures assessing different constructs	85	EPPP Part 2 scores are not well explained by general academic/test taking, personality or intellectual skills, or by variables such as SES	No
Incremental	Improvements in prediction of a criterion variable added by scores on the new test and above other measures	20	Does the EPPP Part 2 predict performance in clinical roles above and beyond the EPPP Part 1	No
Known groups	The extent to which the new measure correctly discriminates between those known to be low and those know	10	People who have received ethical censure from state boards could be paid to take the EPPP Part 2; their scores could	No

Consequences	<p>to be high in a construct The intended and unintended consequences of test use</p>	0	<p>be compared to those with no such record Degree to which EPPP Part 2 shows bias against certain groups of people, including those coming from disadvantaged backgrounds</p>	No
Feasibility	<p>The practical concerns related to use of the new test, such as cost, time, and respondent reactions</p>	0	<p>EPPP Part 2 cost is estimated to be twice the fee of current testing; to require an additional day of testing at a later time point (a protracted process); and will require concentrated preparation from trainees</p>	<p>ASPPB offered a limited period of reduced fee to address concerns raised by graduate trainees and professors</p>

---





*Figure 1.* Relationships between validation of score inference and justification of test use. The solid lines and arrows in the figure represent a linear flow of activities (single-ended arrows) or a recursive process (double-ended arrow). The value considerations (indicated by dashed lines in the upper half of the figure) are not a similarly linear flow; rather, they permeate all of the score inference validation and score use justification process. The solid line from the Results of Test Use box indicates that results directly provide a source of evidence contributing to the corpus of justification evidence, whereas the dotted line from that box indicates that the same results might also produce evidence bearing on the intended score interpretation. [Figure reprinted with permission from Cizek, G. J. (2012). Defining and distinguishing validity: Interpretations of score meaning and justifications of test use. *Psychological Methods*, 17, 31-43. doi: 10.1037/a0026975]