



Association of Real Estate License Law Officials

**Examination Accreditation Program
White Paper on Determining Minimum Passing Scores
Adopted September 16, 2010. Technical revision October 19, 2015.**

The Importance of the Decision

Setting minimum passing scores (cut scores), is among the most important decisions a licensing authority can make. Pass-fail decisions in licensing examinations have significant consequences, both for the candidates who take the examinations and for the public. There are volumes of research and guidance on technical methods for standard setting or recommending cut scores. This paper does not discuss those methods, their application, or their relative merits. Instead, it advises jurisdictional decision makers on the need for a defensible process and provides background on what is generally considered defensible.

Clearly, only the ultimate decision makers for the use of an examination are in a position to establish the passing score—in law, rule, and/or policy. However, the *process* for determining a specific cut score and the *rationale* on which the score is based are critical to the validity of the examination. Validity—the legitimacy of the score interpretation—is cited again and again in professional standards and case law as the most important element when tests are used to make credentialing or employment decisions. For a licensing examination, the interpretation of a passing score is that the person who passes the examination is competent to practice in a manner that protects and does not harm the public. A cut score, then, must be set with a clear concept¹ or definition of who is competent to practice in a manner that protects the public. Further, that concept or definition must take into account specifically the subject matter the examination items cover and how they cover it. If a cut score is set in the absence of such a concept or definition, or there is a failure to agree on the concept or definition, the entire validity of the examination is brought into question. Arbitrary cut scores, even if they are stated in statute, rule, or policy are not defensible in the absence of a professionally acceptable study of the level of performance that should be required of entry-level practitioners.

A Defensible Process

The American National Standards Institute (ANSI) provides guidance for organizations that seek accreditation of their credentialing programs under the international standard ISO 17024: “An assessment can be well designed and built, and closely linked to a thorough practice analysis. Despite satisfying this foundation of test validity, all of this quality and effort is lost if the pass-fail standard is not determined using an acceptable process. The ultimate goal is to maximize the probability that passing candidates are those who deserve to pass, and that failing candidates are those who deserve to fail. These probabilities are maximized only when a criterion-referenced² standard-setting process is applied properly” (American National Standards Institute, 2004, Sec. 4.5).

To be defensible, a standard-setting process that meets professional and legal guidelines must be followed. Such a process should encompass two stages: the recommendation of a cut score by a panel of appropriately qualified experts (judges), and review by decision makers who then establish the cut score to be used. In other words, decision makers should arrive at a cut score only after due consideration of the results of a professionally acceptable standard-setting study. It is true that *all* methods of determining where to set a cut score rely on judgment. However, the *basis* for making those judgments and *who* makes the judgments are critical. *Establishing* the cut score that determines who passes and who fails the jurisdiction’s licensing examinations

ultimately is based on the judgment of the jurisdiction's decision makers. *Recommending* the cut score to the decision makers should be entrusted to professionals who have detailed knowledge and understanding of the knowledge, skills, and abilities (KSAs) required to perform the responsibilities of the job at entry level in a manner that protects the public. Decision makers need to obtain the recommendations of appropriately selected judges that result from application of a formal standard-setting technique (standard-setting study).

Standard-setting judges (raters, panelists) must be knowledgeable about the test content and have experience with the candidate population and the KSAs required for entry-level practice. It is especially important that they have no conflicts of interest, particularly in the outcome of the standard setting. Even when a formal standard-setting study has been conducted in accordance with the most stringent professional standards, the results will be questionable if those who make the judgments are not imminently qualified to do so.

Legal Considerations and Guidance

In the legal framework—laws, regulations, case law—the primary themes that emerge regarding employment testing are job-relatedness, business necessity, and documentation. Although licensing authorities are not technically employers, some laws and guidelines have not been shown to be inapplicable to a state's use of licensing examinations. Measurement professionals, industrial/organizational psychologists, and attorneys consider some of these laws and guidelines in particular, including those cited in the ARELLO® *Guidelines for Accreditation*, to provide important guidance.

The *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014) are considered authoritative guidelines. Among the several standards that apply to setting performance standards or cut scores for examinations, two are particularly important to this paper:

Standard 5.2.1: When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly. (p. 107)
Standard 11.16: The level of performance required for passing a credentialing test should depend on the knowledge and skills necessary for credential-worthy performance in the occupation or profession and should not be adjusted to control the number or proportion of persons passing the test. (p. 182)

The *Uniform Guidelines on Employee Selection Procedures*, adopted in 1978 by the EEOC, U.S. Civil Service Commission, Department of Labor, and Department of Justice, apply to all selection procedures used as a basis for any employment decision and are given deference by the courts. The Uniform Guidelines outline standards for selection procedures, including: (1) establishing validity (job analysis and other steps), (2) use of selection procedures, (3) setting cut scores, and (4) documentation.

In "Legal Issues for Licensure and Certification," Carson (2001) thoroughly reviews legal issues in licensure and certification testing. She concludes that "Although licensing and certifying examinations have generally withstood judicial scrutiny, it is in the best interest of a licensing authority or certifying entity and the public it serves to have as much evidence as possible available to support its standard-setting judgment" (p. 442). Underscored throughout the analyses is the importance of validity, appropriate standard-setting procedures, and documentation.

ARELLO® Accreditation Requirements

Examination providers with programs accredited under the *ARELLO® Examination Accreditation Program* are required to conduct standard-setting studies in accordance with professional standards. ARELLO® provides specific direction on the obligations of examination providers with respect to cut scores on examinations accredited under the *ARELLO® Examination Accreditation Program*. To be consistent with the *ARELLO® Guidelines for Accreditation*: The examination provider shall recommend a minimum passing score (cut score) derived using a method that is designed to be nationally representative and consistent with the *Standards for Educational and Psychological Testing*, the *Uniform Guidelines on Employee Selection Procedures*, the *Civil Rights Act of 1991*, and other regulations and standards applicable in the jurisdiction(s) in which the examination is to be used. When a jurisdiction does not adopt the recommended cut score, the examination provider shall advise that jurisdiction that the cut score is a critical element in examination validity and that arbitrary cut scores set merely by statute, rule, or policy are not defensible in the absence of a professionally acceptable study of the level of performance that should be required of entry-level practitioners. (p. 7)

Using the results of an accredited provider's standard-setting study provides a measure of credibility. Examination professionals design and conduct the studies, involving experts from a variety of jurisdictions, and carefully document the process and results. Decision-makers in the jurisdictions that rely on these examinations to make licensing decisions must, however, use the valuable information provided in an informed, reasoned manner in the final decision.

Considerations for Decision-Makers

Geisinger and McCormick (2010) point out that there are legitimate reasons for adjusting recommendations to arrive at an adopted cut score point and that decision makers need to use “the input of the standard-setting panelists as one piece of information among multiple sources” (p. 38). Some of the considerations they discuss that may be applicable to jurisdictions in determining whether to accept the recommendations of an examination provider or to adopt a different cut score are previous standards and historical pass rates, political feasibility, relative costs of misclassifying candidates (failing someone who should be licensed and passing someone who should not), adverse impact (a substantially different rate of selection that works to the disadvantage of members of a race, gender, or ethnic group), results from different standard-setting sessions or techniques, and others. They urge attention to anomalies (unusual or unexpected occurrences) in the standard-setting rating process.

Ultimately, the most critical factor in considering adopting or altering cut scores recommended on the basis of a thorough passing-score study is the soundness and defensibility of the rationale. Decision makers must weigh heavily the results of deliberate, data-based, criterion-referenced formal processes and documentation of evidence supporting a recommendation. Such processes and documentation should be designed and conducted by measurement professionals with appropriately selected and trained judges who have detailed knowledge and understanding of the KSAs required to perform the responsibilities of the job at entry level in a manner that protects the public. If a different cut score is adopted in the absence of a defensible rationale, the entire validity of the examination is brought into question.

Recommendations

In summary, jurisdictions should employ a two-stage process in determining cut scores for licensing examinations: the recommendation of a cut score by a panel of appropriately qualified experts, and review by decision makers who then establish the cut score to be used. The process should include a standard-setting study conducted by professionals who have the educational

background, knowledge, and skills to do so according to professionally accepted standards. It is important that these professionals observe panelists carefully and pay attention to the data they collect to be sure that the panelists understand their task, that they make their ratings independently, and that individuals' ratings do not appear to be inappropriate. It is imperative that they carefully document the process and all of these observations and report them straightforwardly to jurisdictional decision makers. It is then incumbent upon jurisdictional decision makers to review all aspects of the study: who the panelists were and how they were selected (qualifications, recognition in the field, absence of conflict of interest, ability to understand the task, representativeness), how the study was conducted (design of study, materials, instructions to panelists, conditions, focused attention to the task, the data presented), what unexpected events may have occurred, and how the results were documented and analyzed. Reviewers may wish to see that panelists have signed an affidavit attesting to their lack of conflict of interest, understanding of the subject matter, understanding of the KSAs of an entry-level practitioner, understanding of the rating task, and other factors the reviewers consider important.

References

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

American National Standards Institute (2004). *Guidance on psychometric requirements for ANSI accreditation*. New York, NY: Author.

Carlson, J. V. (2010) *ARELLO® Guidelines for Accreditation* (11th ed.). Littleton, CO: Association of Real Estate License Law Officials.

Carson, J. D. (2001). Legal issues in standard setting for licensure and certification. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 427-444). Mahwah, NJ: Lawrence Erlbaum.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice (1978). Uniform guidelines on employee selection procedures. *Federal Register*, 43(166), 38290–38315.

Geisinger, K. F., & McCormick, C. M. (2010). Adopting Cut Scores: Post-Standard-setting panel considerations for decision makers. *Educational Measurement: Issues and Practice*, 29(1), pp. 38-44.

Sireci, S. G., & Hambleton, R. K. (2009). Mission: Protect the public: Licensure and certification testing in the 21st century. In R. Phelps (Ed.). *Correcting fallacies about educational and psychological testing*. Washington, DC: American Psychological Association.³

¹ A concept is a mental “picture” combining all characteristics.

² Criterion-referenced means based on a specific criterion or standard related to the subject matter rather than a relative one, such as comparing individuals to one another.

³ In addition to the references cited that deal specifically with the issue of cut scores, see this chapter for a simple, broad discussion of current considerations, trends, and issues in testing for credentialing.

Author: JoEllen V. Carlson, PhD

END of White Paper. Additional recommendations follow.

Following are examples of replacement language for provisions of statute, rule, or policy that may reference an arbitrary or undocumented percentage or specific numerical passing score, allowing jurisdictions to adopt a passing score that has been derived from a professionally acceptable study of the level of performance required of entry-level practitioners:

- Passing scores shall be derived using a method that is designed to be nationally representative and consistent with the *Standards for Educational and Psychological Testing*, the *Uniform Guidelines on Employee Selection Procedures*, the *Civil Rights Act of 1991*, and other regulations and standards applicable in *(name of the jurisdiction)*.
- Candidates for licensure must take a Commission approved examination and achieve a passing score as determined by psychometrically sound criterion-related methods associated with assessment of minimal competence.
- To pass the real estate licensing examination, an applicant must attain a score at least equal to the passing score established by the Commission in compliance with psychometric standards for establishing passing scores for occupational licensing examinations as set forth in the *Standards for Educational and Psychological Testing* jointly promulgated by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education.