

Using Performance Assessments to Determine Competence in Clinical Athletic Training Education: How Valid Are Our Assessments?

Gayle A. Thompson, PhD*; Robert Moss, PhD†; Brooks Applegate, PhD*

*Western Michigan University, Kalamazoo; †Department of Kinesiology, Albion College, MI

Context: Validity arguments can be used to provide evidence that instructors are drawing accurate conclusions from the results of students' clinical performance assessments (PAs). Little research has been conducted in athletic training education to determine if the evidence supports the use of current PAs. Measurement theories designed to provide this evidence can be confusing and unfamiliar to athletic training educators.

Objective: The purpose of this article is to present contemporary concepts of validity and suggest approaches athletic training educators can use to offer evidence to support the best assessment methods.

Background: Educators often use PAs to determine a student's competence for professional practice. Competence is a complex concept that is difficult to define clearly, thus making assessments of competent performance difficult as well. Most methods of PA used in athletic training education can be classified into 2 general approaches: behavioral and holistic. Athletic training educators, in an attempt to develop effective, appropriate, and user-friendly PAs to evaluate students, may be measuring skill but not truly measuring competence.

Description: Modern validity concepts focus on the interpretations and meanings of assessment scores, not just on the characteristics of the test itself. Using an updated concept of validity can guide the development of competence PAs to determine if educational outcomes are being met. A framework for developing a validity argument is presented.

Conclusions: Validity can be used to provide a simple, but rational, defense of what clinical educators do. Knowing the process of establishing validity evidence will help educators revise PAs and educational standards to further promote the profession.

Key Words: Validity, behaviorial assessment, holistic assessment

Dr Thompson is Program Director for the Undergraduate Athletic Training Program in the Human Performance and Health Education Department at Western Michigan University. Please address all correspondence to Gayle A. Thompson, PhD, Western Michigan University, HPHE, 1903 W Michigan Avenue, Kalamazoo, MI 49008-5426. gayle.thompson@wmich.edu.

Full Citation:

Thompson GA, Moss R, Applegate B. Using performance assessments to determine competence in clinical athletic training education: how valid are our assessments? *Athl Train Educ J*. 2014;9(3):135–141.

Using Performance Assessments to Determine Competence in Clinical Athletic Training Education: How Valid Are Our Assessments?

Gayle A. Thompson, PhD; Robert Moss, PhD; Brooks Applegate, PhD

INTRODUCTION

Athletic training educators strive for evidence that promotes the best practices in educational techniques and assessments. Yet methods for finding this sort of evidence can remain elusive. Of course, students' performances indicate some level of evidence. If students perform appropriately in a given situation, perhaps it is because of the educational approach; however, establishing what is appropriate in a given situation can be challenging in the health care realm because there is often more than one situational response. At times, the tools or performance assessments (PAs) educators use to determine competence may fail to take into account problem solving and decision making because checklist-style assessments are easier to standardize and grade.¹ However, athletic training educators want quality assessments of students' performances that require students to use critical thinking and demonstrate the ability to make sound professional decisions. Therefore, PAs should also measure decision-making skills. Little research has been conducted in athletic training education, however, to determine if evidence supports the use of current PAs. Measurement theories designed to provide this evidence can be confusing and unfamiliar to athletic training educators. The purpose of this article, therefore, is to present contemporary concepts of validity and suggest approaches athletic training educators can use to offer evidence to support the best forms of assessment. Although validity is rarely the most engaging topic for athletic training educators, fully understanding the concept of validity can lead to improved methods of evaluating students' clinical abilities and provide evidence to defend current practices.²

BACKGROUND

Performance assessment is a broad term that essentially describes most forms of educational appraisal that measure a student's ability to perform a task.² Primarily, PAs involve direct appraisal of a task or behavior in a realistic situation that allows for inferences based on application of knowledge.² Like other methods of educational testing, PAs are useful for determining students' strengths and weaknesses, areas of achievement, and areas that require intervention.³ The emphasis of PA in a real-life context is often thought to improve the fidelity or authenticity of the evaluation, or more accurately, the inferences of the PA.² Therefore, educators often use PA to determine a student's competence for professional practice.⁴

A variety of PA styles are used in athletic training education, depending on the desired educational outcomes. At the introductory level, a checklist approach is very useful because the educational goals are simply to determine if a student can perform a given task. However, as students become more advanced in their education, it is desirable to determine if they are making the appropriate decisions regarding when to perform certain tasks. In other words, decision making and competence are being evaluated in a given situation.

Competence is a complex concept that is difficult to define clearly because even among experts in a particular profession there are a variety of approaches to handling situations that arise in professional practice. This, in turn, makes assessments of competent performance difficult as well because it is challenging to clearly define how an examinee should respond to a given situation during an assessment. However, most professionals have a very good idea of what competence and, even more so, incompetence look like. They also have a clear idea of the expectations of professional practice.⁴ A variety of definitions of competence exist in the literature, each with its own nuances.^{1,4-7} For this article we will use a definition of competence that incorporates the relationship between performance and determinations of competence. Competence can be defined as the "degree to which the individual can use the knowledge, skills, and judgments associated with the profession to perform effectively in the domain of possible encounters defining the scope of professional practice."⁴(§166) The challenge lies in finding appropriate ways to accurately draw conclusions regarding competence from a PA.

Allied health educators use a variety of PA methods to determine a student's level of competence. These methods range from direct observation by a clinical mentor while working alongside a student to formally structured practical PA that involves executing specific skills or applying decision making to orchestrated scenarios. Performance assessments are an appropriate way to determine the performance of psychomotor tasks necessary for professional practice.⁸

Most methods of PA used in athletic training education can be classified into 2 general approaches: behavioral and holistic.¹ As previously indicated, both approaches are useful in clinical education depending on the specific educational outcomes desired.⁹ The behavioral approach stems from the belief that competence can be demonstrated or directly observed through students' performances on specified tasks.¹ A primary characteristic of the behavioral approach is the standardization of test items and criteria for performance into discrete elements. Performance expectations are clearly delineated so that there is no discrepancy on what constitutes a satisfactory performance.¹ This aspect of the behavioral approach is what makes it so appealing. By standardizing the assessment, proponents assume that the assessment is unbiased and objective; therefore, an adequate performance on the assessment indicates a student's level of competency.⁶ In allied health, this approach is seen in any variety of structured clinical examinations or in parochial checklists as criteria for PA.¹⁰⁻¹³ Individualized psychomotor tasks are given to a student to perform, and an assessor has a list of specifications or outcomes that are marked as satisfactory or unsatisfactory based on the student's performance. Many will remember the old practical section of what is now known as the Board of Certification (BOC) exam as an example of this. In athletic training clinical education, basic psychomotor skills are often assessed using the behavioral approach.¹⁴

One of the main criticisms of the behavioral approach is that atomizing skills into discrete tasks ignores decision making, professional intelligence, and the complexity of performance in real-life situations.^{1,6,15,16} Fragmented tasks are seen as adding up to less than the whole, as they do not adequately represent the scope of practice and ignore the importance of decision making in competent performance.⁶ Another concern is that there is no guarantee of generalizability of the performance to other contexts, thereby limiting the ability to truly determine competency.^{15,17}

In contrast to the behavioral approach is the holistic approach to PA, which integrates professional judgment and critical thinking into PA and acknowledges that there may be more than one way of practicing competently within a given situation. The foundation for determining competence is a knowledge base that allows an individual to use professional judgment in the appropriate context using personal attributes such as critical thinking and the ability to communicate.¹ Admittedly, the approach lacks the simplicity and clarity of the behavioral approach as there are often no straightforward criteria outlining an acceptable performance or how it can be evaluated.¹ However, the holistic approach calls for PAs that incorporate evaluation of a number of elements of competence simultaneously in a realistic, relevant environment. In allied health, this approach is most commonly seen in portfolios, certain practical simulations, direct observations, and problem-based learning scenarios.¹⁸

Clinical education in allied health relies on many forms of PA to make decisions on the level of competence of students. Because it is impossible to evaluate students on their abilities over the entire spectrum of professional practice, judgments are made based on a limited number of observations, which may not be a representative sample or done under conditions that mimic actual practice.⁴ Therefore, validation of measures of professional competence is crucial.

VALIDITY

Although they may understand the importance of validity in educational assessment, many athletic training educators hold a traditional, outdated view of validity.^{9,19–23} The concept of validity has evolved from the definition “that a test measures what it claims to measure”^{19,20} to “refer(ring) to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of a test.”²³⁽⁸⁹⁾ The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretation.²⁴ Parameters of the test are still important to establish validity, but it is incorrect to state that a test itself is valid or not valid.³ A PA can only be considered valid to the extent that evidence can be provided to support the interpretations or meanings attributed to the assessment.²³ Therefore, a test cannot dichotomously claim to be a good test or a bad test; it is dependent on the evidence provided to justify the use of a test and how the results will be interpreted.^{2,21,23,24}

Recent athletic training research demonstrates the desire to address validity to confirm that educators “are properly measuring program outcomes.”^{20(§37),25} However, it is important to ensure that the most accurate representations of the concept are being used. Despite its general adoption by many medical educators,^{4,8,21,26–28} most athletic training

research has not adopted this more accurate concept of validity that shifts the focus to the interpretations and meanings assigned to the test results instead of the qualities of the test itself. Measurement experts argue that if validity has no meaning beyond the content of a list of test items, educators are failing to consider the ways tests are used in practice.^{22,29} Consequently, assessments are only useful when used to make decisions and determine our outcomes. Whether decisions are on competence, educational practice, or program outcomes, they must be justified. Appropriate use of validity theory can guide this pursuit to justify and determine best practice.

In addition, references to validity often erroneously imply that one could arbitrarily choose from any multitude of equally legitimate types of validity, the most popular being construct, content, and criterion. It is more accurate to say these are types of evidence that contribute to establishing validity. There are not different kinds of validity, only various types of evidence to support a test’s intended use or interpretation.²¹ Essentially, different types of evidence can be used to clarify test interpretations in a particular contexts.⁹ If the context changes, so must the evidence.

Any discussion of validity must also include reliability. Reliability, or consistency, was traditionally considered a precursor to validity in that a test could not be valid if it was not first consistent.³⁰ However, changing perspectives on validity often include reliability estimations as an important component of validity evidence, instead of the two being totally separate entities.^{21,31} These two measurement concepts must work synergistically to provide evidence of a value of a measurement, especially in regard to the ability to generalize or extrapolate the observations of the measurement to a wider domain.

In athletic training education, standardized scoring criteria may be used at times in an attempt to improve the consistency or reliability of PA, even when the behavioral approach may not best meet the needs of educational outcomes. Standardization of conditions to improve consistency may also risk narrowing the ability to extrapolate the results beyond compatible conditions, thereby risking the fidelity, or authenticity, of the PA as well as the generalizability.² Practicing professionals also recognize the dilemma of sacrificing the ability to apply skills in a real-life context in order to standardize a PA.^{1,2,30} Standardized criteria also do not take into account that there may be more than one correct way to competently respond to a given problem.⁶ Educators want to use quality assessments but perhaps are unsure how to provide evidence to support the use of more holistic, open-ended forms of PA. Standardization also protects educators from the threat of rater bias in PA. There is often concern that educators may rate a performance at a gut level, allowing personal feelings regarding the student or variations in a teacher’s mood to affect the grade a student receives.³² This potential for subjectivity is viewed as a threat to validity and reliability, and many researchers recognize it as a limitation of PA.^{2,6,8,33,34} Yet some researchers argue that ignoring the expertise of raters and instructors by attempting to standardize, and therefore removing individual expertise in scoring PA, is also a threat to the validity of the assessment.^{16,26,35} Including some aspects of subjective assessment may speak to the art of health care that extends beyond what can be

captured on a standardized scale.²⁶ Although issues of subjectivity are inherent with some PAs and should be addressed, they do not necessarily negate the evidence to support their use.^{16,35} It should be emphasized that clinical educators must be well trained in order to develop the expertise to effectively evaluate students.

Another safety net athletic training educators may have relied on too much is using fairly simple measures of reliability in an effort to defend the value of an assessment. Statistical measurements of inter-rater reliability offer empirical data about rater reliability and are relatively simple to calculate and interpret. Studies have found that inter-rater reliability can be brought to acceptable levels with training of the raters, standardized test administration procedures, and clear scoring rubrics to address concerns regarding reliability.^{6,30} As previously stated, evidence to support the reliability of a test is closely linked with being able to provide validity evidence.^{4,21,30,31} Therefore, simple calculations like inter-rater reliability, while offering valuable information about an assessment, may be over relied upon in an attempt to provide some evidence of the value of a measurement in light of the ambiguity surrounding validation of PAs.³³ However, relying solely on these methods to address reliability, and therefore validity, often does not go far enough to provide validity evidence that justifies the use of a given PA.

VALIDITY AND PERFORMANCE ASSESSMENTS

The behavioral approach to PA is an uncomplicated way to measure educational outcomes with what appear to be sound principles of measurement. Athletic training educators, in an attempt to develop effective, appropriate, and user-friendly PAs to evaluate students, may be measuring skill but not competence. Critics of the behavioral approach claim it decontextualizes skills and ignores the complexity of real-world situations.¹ As such, the behavioral approach may assess an isolated activity but would not be an indicator of overall competent practice.¹⁶ Therefore, inferences made based on the behavioral approach would not adequately support the construct of competence. Other experts counter that there is a place in education for breaking down psychomotor skills into specific tasks.³⁴ Perhaps the dispute is over the intended use of the PA. If the goal of PA is to determine performance of a specific psychomotor skill, assessing decontextualized tasks may achieve this goal. On the other hand, if the intended use of PA is to determine competence, then the behavioral approach may fall considerably short of the goal and threaten its validity. Yet more holistic PAs have limitations as well, primarily the challenge in proving traditional concepts of validity and reliability, as previously described. Despite genuine interest in holistic forms of PA, such as problem-based learning, athletic training educators are still struggling with the implementation of such techniques.¹⁸

Several forms of PA have a place in athletic training education, depending on the desired educational outcomes. Therefore, addressing the validity of PA to assess competence may involve negotiating the strengths and weaknesses of the evidence when interpreting the scores.⁴ This evidence must also take into account the limitations of the inferences. It may be that although a certain PA does not ideally meet each validity criteria, it is still the best measurement tool available

for assessments of competence based on desired educational outcomes.

So educators are left with presenting the appropriate evidence to systematically support the validity of PAs of competence, regardless of the approach. Behavioral or holistic, what types of evidence are considered relevant to examine the overall validity of PAs of competence? Systematically investigating the validity of PAs is the best way to defend current practice. Developing a practical framework for providing useful validity evidence that supports educational goals will make PAs of competence more effective and improve educational practice.

RECOMMENDATIONS FOR EDUCATIONAL PRACTICE

Establishing Validity Evidence

The first obstacle to overcome when establishing validity is ensuring that educators are given an accurate and comprehensive concept definition. The next step is to develop a user-friendly method to provide evidence, or a validity argument, to support the use of PA. Many approaches have been offered, each building on the validity concepts previously discussed, which essentially incorporate many aspects of scientific inquiry.^{2,4,9,21,23,27,28,34} The framework introduced here was developed some time ago for PA of competence in allied health, but it is a foundational approach to defend validity used by medical educators.^{4,27,28} This framework offers an easily understood, practical approach to determining validity.

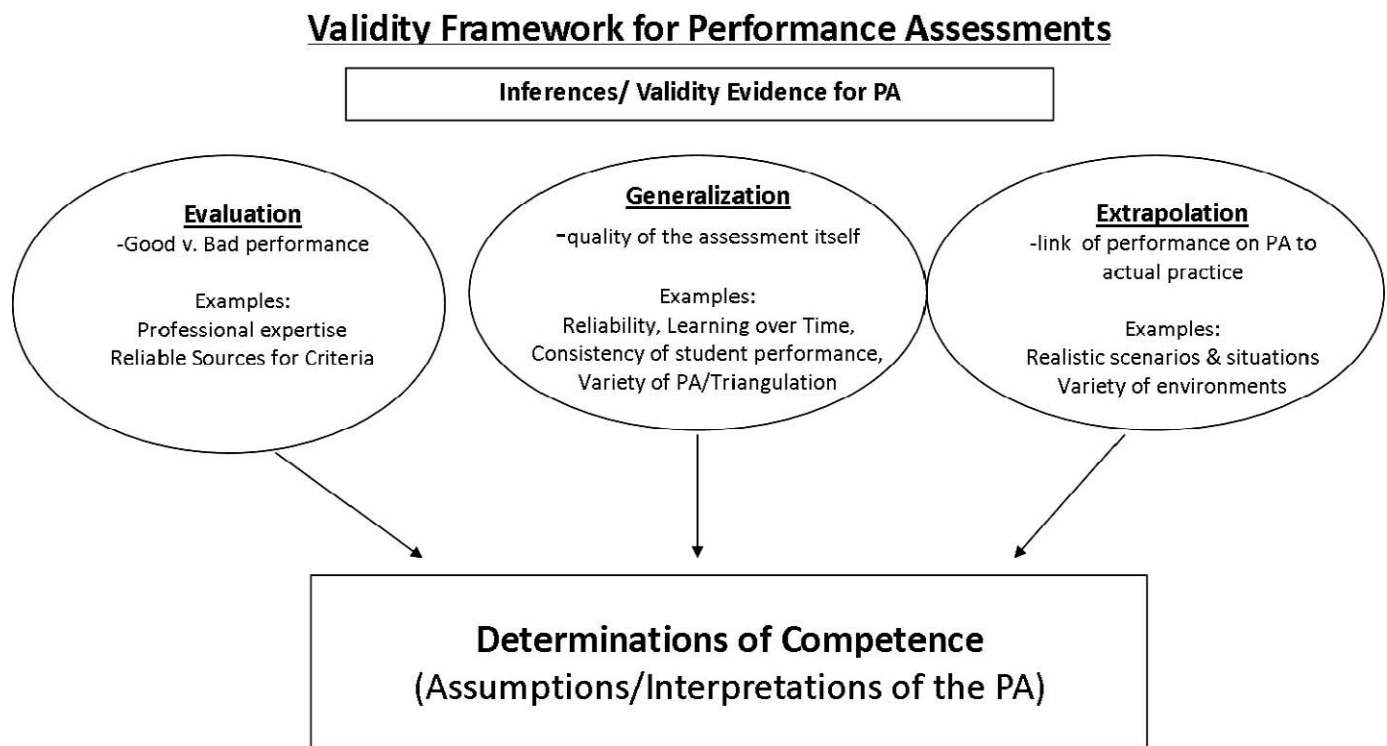
Validity Framework for Assessments of Competence

This approach begins with development of specific statements for the intended use of a PA, then identifies the strengths and weaknesses of these inferences (Figure).^{4,29} All forms of assessment have strengths and weaknesses, and the intent of the framework is to identify them so that rational decisions can be made to defend interpretations of PA scores. Therefore, validity is established by determining how credible and reasonable the link is between the PA and determinations of competence.⁴ The approach emphasizes defining the limitations in the interpretations and addressing them in order to strengthen the arguments that support the conclusions of PA.

The framework involves 3 primary inferences, or aspects of evidence: evaluation, generalization and extrapolation.⁴ Each of these can easily be applied to PAs used in entry-level athletic training clinical education and will be further expanded on in the sections that follow.

Evaluation. Evaluation involves the criteria that are established to determine the quality of the responses, delimiting a good performance from a bad one.⁴ This can also be referred to as *content evidence*.²⁷ Obviously, in order to ascertain that a student is competent, some guidelines must be offered as a basis for decision making on the performance. For example, to create a PA to assess a student's ability to perform the Ottawa ankle rules the developer would rely on accepted texts and expert opinion to establish the criteria for the task to be successfully completed, that is, palpation at the proper landmarks and patient function.³⁶ This supports the content, or evaluation, aspect of the assessment as long as the goal of

Figure. Validity framework for performance assessments (PA).



the PA is simply to determine that a student can perform the essential tasks of the rule. This PA would not, however, provide evidence to determine that the student is competent in diagnosing ankle injuries because the content only addresses one aspect of ankle assessment. The developer must also be sure to defend the soundness of the sources of the criteria, especially in instances where opinions vary for a particular task.

Generalization. Generalization refers to the internal structure of PA.⁴ This inference is based in generalizability theory, a measurement theory that examines how well a one-time observed performance represents how a student might potentially perform the same task in all possible circumstances.³⁷ Any given performance on a PA is only an estimate from a wide range of possible ways and circumstances in which a student may perform that task. Therefore, it is essential to consider the limitations of making too broad an assumption regarding competence based on limited assessment opportunities. This is also where evidence of reliability and consistency of assessment conditions will need to be addressed. The procedures of a PA should be defensible.

In practice, generalization can be argued by evidence of learning over time, an athletic training education concept that emphasizes multiple assessments of the same skill set.¹⁴ For example, evidence to support the inference of generalizability could be provided when students demonstrate the ability to perform the Ottawa ankle rules in a variety of appropriate contexts. A typical progression would be from a didactic approach to a classroom lab to an actual ankle assessment supervised by a preceptor. The PAs in each of these environments offer increasing support for the application of the skill in a variety of situations. By providing several opportunities for students to demonstrate skill and competence, using a variety of methods, clinical educators are more

likely to gain a comprehensive picture of students' true abilities with less variation between performances.

Extrapolation. The third inference, extrapolation, directly refers to the ability of performance during the assessment to be linked to performance in actual practice.⁴ Essentially, the interpretations of performance on PAs are more likely to be defensible in highly realistic or authentic conditions. Educators see this exhibited when students successfully perform a skill in a well-controlled classroom lab but struggle the first time a preceptor encourages them to perform the task on a real patient. Educators strive to create testing environments that are authentically related to actual professional practice, yet establishing quality, defensible measurements in these more authentic situations is challenging. Fidelity of an assessment cannot be assumed by simply observing students in actual practice. The evaluator's presence could influence the student's performance, for better or worse, and thus, still involves some level of extrapolation.⁴ Consequently, multiple PAs in a variety of realistic situations will offer more defensible validity evidence for extrapolation.

Evidence of extrapolation for the Ottawa ankle rules example could include correlating the scores of multiple PAs of the skill. A behavioral PA of the rules could be correlated at a later time with a more holistic PA of a complete ankle evaluation in a classroom lab setting. Both simulated conditions could then be correlated with an assessment of the skill provided by a student's preceptor when applied and performed during the course of a clinical rotation. Ultimately, the results of a student's PA should be confirmed by successful completion of the BOC examination and subsequent performance in professional practice. Clearly, the long-term nature of this evidence fails to provide formative support for the use of the PA.

There are a few general considerations for implementing this framework. Although evidence should be presented for all 3 inferences (ie, evaluation, generalization, and extrapolation), there may be situations where some inferences should be given more or less priority when selecting an appropriate PA method.⁴ For example, if high-quality evidence exists to support the use of a given technique, as with the example of the Ottawa ankle rules, then additional evaluation evidence would not likely be necessary to support its inclusion in a PA. However, evidence of generalization and extrapolation are not as inherent and would need to be emphasized depending on the conditions and context of the PA. This could include evidence that those administering the PA are highly skilled in the use of the Ottawa ankle rules and that performance is assessed more than once in a variety of settings. The extent of the evidence required may vary depending on the consequences of the PA, the criticality of the assessment, and the expertise of the educator.²⁷

Although this framework specifically addresses the use of PA as it relates to interpretations of competence, other types of validity evidence can and should be provided when the PA will be used to make other types of decisions. Although not addressed within the context of this article, other authors describe internal or measurement evidence, consequential evidence, and process evidence for decisions on the appropriateness of the validity argument.^{27,28}

It is essential to carefully consider the goals of PA and the interpretations to be made in order to guide *any* validity argument. Regardless, the emphasis should be on discovering the weaknesses in the argument to support the use of a PA and being able to plausibly address those weaknesses.⁴ It is important to emphasize that determining validity is not a dichotomy; tests are not either valid or invalid. There can be varying degrees of how well the validity evidence provided for a PA supports the interpretations to be drawn from its use.^{23,38} Finally, using a variety of assessments to determine competence also increases the validity of the interpretations.⁶ Using multiple measures to support or refute interpretations of PA can dramatically strengthen a validity argument.

CONCLUSIONS

Often terms like *validity* in the field of athletic training are met with trepidation. However, validity does not have to be imposing; it involves a simple, but rational, argument or defense of what clinical educators do. Knowing the process of establishing validity evidence will help educators revise PAs and educational standards to further promote the profession. It may also help with the ambiguity over implementing the latest edition of the *Athletic Training Educational Competencies*.³⁹ Clinical educators should not hesitate to use their unique professional abilities to determine their students' competence.

The primary intent of this article is to empower clinical educators to use validity to analyze and defend their current use of PA in their current practice. It is time for athletic training educators to either use the tools that will reflect best educational practices or discover that practices need to be revised. Are PAs truly determining how competent our students are when they enter the profession? If so, we need to share that information. If not, we need to do better.

REFERENCES

1. Goncz A. Competency based assessment in the professions in Australia. *Assess Educ Princ Policy Pract*. 1994;1(1):27–42.
2. Kane MT, Crooks T, Cohen A. Validating measures of performance. *Educ Meas Issues Pract*. 1999;18(2):5–17.
3. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 1999.
4. Kane MT. The assessment of professional competence. *Eval Health Prof*. 1992;15(2):163–182.
5. Cross V, Hicks C, Barwell F. Comparing the importance of clinical competence criteria across specialties: impact on undergraduate assessment. *Physiotherapy*. 2001;87(7):351–367.
6. McMullan J. Competence and its assessment: a review of the literature. *Br J Podiatry*. 2005;8(2):49–52.
7. Milligan, F. Defining and assessing competence: the distraction of outcomes and the importance of educational process. *Nurs Educ Today*. 1998; 18(4):273–280.
8. Watson R, Stimpson A, Topping A, Porock D. Clinical competence assessment in nursing: a systematic review of the literature. *J Adv Nurs*. 2002;39(5):421–431.
9. Messick S. Standards of validity and the validity of standards in performance assessment. *Educ Meas Issues Pract*. 1995;14(4):5–8.
10. Task force for the development of student clinical performance instruments. The development and testing of APTA clinical performance instruments. *Phys Ther*. 2002;82(4):329–353.
11. Bondy KN. Clinical evaluation of student performance: the effects of criteria on accuracy and reliability. *Res Nurs Health*. 1984;7(1):25–33.
12. Gregg A, Dehn R. An analysis of clinical evaluation methods in physician assistant education. *Perspect Physician Assist Educ*. 1999;10(4):185–193.
13. Hanley E, Higgins A. Assessment of clinical practice in intensive care: a review of the literature. *Intensive Crit Care Nurs*. 2005; 21(5):268–275.
14. National Athletic Trainers' Association. *Athletic Training Educational Competencies*. 4th ed. Dallas, TX: National Athletic Trainers' Association; 2006.
15. Huddle TS, Heudebert GR. Taking apart the art: The risk of anatomizing clinical competence. *Acad Med*. 2007;82(6):536–541.
16. ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med*. 2007;82(6):542–547.
17. Chambers MA. Some issues in the assessment of clinical practice: a review of the literature. *J Clin Nurs*. 1998;7(3):201–208.
18. Heinrichs KI. Problem-based learning in entry-level athletic training professional-education programs: a model for developing critical-thinking and decision-making skills. *J Athl Train*. 2002;37(4 suppl):S-189–S-198.
19. Arnold B, Gansneder BM, Perrin DH. *Research Methods in Athletic Training*. Philadelphia, PA: FA Davis; 2004.
20. Eberman L, Kahanov L. Defining outcomes and creating assessment tools for AT education, part 2. *Int J Athl Ther Train*. 2011;16(1):37–41.
21. Goodwin LD. Changing conceptions of measurement validity. *J Nurs Educ*. 1997;36(3):102–107.

22. Moss PA. Reconstructing validity. *Educ Res.* 2007;36(8):470–476.
23. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ.* 2003;37(9):830–837.
24. Kane MT. Current concerns in validity theory. *J Educ Meas.* 2001;38(4):319–342.
25. Eberman L, Kahanov L. Defining outcomes and creating assessment tools for AT education, part 1. *Athl Ther Today.* 2010;15(6):41–44.
26. Mahara MS. A perspective on clinical evaluation in nursing education. *J Adv Nurs.* 1998;28(6):1339–1346.
27. Andreatta PB, Marzano DA, Curran DS. Validity: what does it mean for competency-based assessment in obstetrics and gynecology? *Am J Obstet Gynecol.* 2011;204:384.e1–e6.
28. Andreatta PB, Gruppen LD. Conceptualising and classifying validity evidence for simulation. *Med Educ.* 2009; 43:1028–1035.
29. Kane MT. Terminology, emphasis, and utility in validation. *Educ Res.* 2008;37(2):76–82.
30. Moss PA. Can there be validity without reliability? *Educ Res.* 1994;23(2):5–12.
31. Hurley WL, Denegar CR, Hertel J. *Research Methods: A Framework for Evidence-Based Clinical Practice.* Philadelphia, PA: Lippincott Williams & Wilkins; 2011.
32. Mehrens WA, Lehmann IJ. *Measurement and Evaluation in Education and Psychology.* 4th ed. Orlando, FL: Harcourt Brace College Publishers; 1991.
33. Linn RL, Baker EL, Dunbar SB. Complex, performance-based assessment: expectations and validation criteria. *Educ Res.* 1991; 20(8):15–21.
34. Messick S. The interplay of evidence and consequences in the validation of performance assessments. *Educ Res.* 1994;23(2):13–23.
35. Govaerts MJ, van der Vleuten CP, Schuwirth LW, Muijtjens AM. Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Adv Health Sci Educ Theory Pract.* 2007;12(2):239–260.
36. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Annu Emerg Med.* 1992;21(4):384–390.
37. Crocker L, Algina J. *Introduction to Classical and Modern Test Theory.* Fort Worth, TX: Harcourt Brace Jovanovich College Publishers; 1986.
38. Denegar CR, Fraser M. How useful are physical examination procedures? Understanding and applying likelihood ratios. *J Athl Train.* 2006;41(4):201–206.
39. National Athletic Trainers' Association. *Athletic Training Educational Competencies.* 5th ed. Dallas, TX: National Athletic Trainers' Association; 2011.