Educational Measurement
ISSUES AND PRACTICE

NCME national council on measurement in education

Educational Measurement: Issues and Practice
Winter 2014, Vol. 33, No. 4, pp. 25–28

Formative Assessment and Test Security: The Revised *Standards* Are Mostly Fine; Our Practices Are Not

Steve Ferrara, *Pearson*

In response to the editor's charge, I have chosen to write about two issues that the *Standards*—and the educational testing community—have not addressed adequately. First, as Plake and Wise (2014) indicate, the *Standards* do not address classroom formative assessment practices or instruments. They tell us that the current joint committee adopted the previous committee's decision not to address classroom assessment in the *Standards*. I agree with the committee's decision, because of the target audience for the *Standards*: "[measurement] students, policy-makers, and test users as well as measurement professionals" (p. 6). Beyond the *Standards*, I think that we need to take action on classroom assessment practices. Second, while the *Standards* address prevention and detection of breaches of test security, I believe that the testing and policy communities do not provide adequate guidance on investigating and resolving violations of test security. Both of these inadequacies are relevant to my charge from the editor: to address the tension in commercial testing contexts between the business mission of meeting market demands and the professional demand of adhering to and promoting professional standards and practices. I hope that you will agree that we should address these inadequacies, even if you disagree with the actions that I propose.

Before I elaborate on these recommendations, a disclosure is necessary. I served as NCME's liaison to the joint committee. References to my reports to the NCME Board of Directors in this capacity appear in Board minutes during 2009 to 2013. My reports focused on the deliberation and revision process, not the substance of the revisions, and I tried to identify issues that the Board should be aware of (e.g., the significant reorganization of standards on fairness in testing). By agreement with the chairs of the joint committee, I participated in discussions judiciously, made occasional comments to add an idea to a discussion, and had no involvement in decision making. My measured participation was a minor influence in the committee. Even with all of that, I have to acknowledge that my views on the joint committee and the revised *Standards* are not completely objective. Several times at NCME meetings I have expressed my admiration for the deliberation and revision process and support for the revised *Standards*. The Plake and Wise description of the process is an accurate and fair representation of what happened in the joint committee.

Steve Ferrara, Pearson, Research and Innovation Network, Pearson School, c/o 1242 C Street, SE, Washington, DC 20003; steve.ferrara@pearson.com

Classroom Assessment Practices and Embedded Formative Assessment Activities: Assessment *for* Learning

Several of our colleagues have defined classroom assessment practices as any assessment activity that provides information about student learning that students and teachers can use, in a timely fashion, to decide about what to teach and learn next and to engage students in assessment of their own learning (e.g., Cizek, 2010). The popular term is "assessment *for* learning" (Stiggins, 2005). These formative assessment activities can include teacher oral questioning, teacher-made tests, commercially provided formative assessment instruments that are aligned to state content standards, and assessment activities embedded in commercial instructional content and materials. Schools and school systems want help with formative assessment; educational publishers provide formative assessment products and professional development services. However, even training thousands of teachers per year barely makes a dent in the 3.3 million K-12 public school teachers in the United States in 2012 (see http://nces.ed.gov/fastfacts/display.asp?id=28). Quite often, the formative assessment instruments publishers provide are not embedded in instructional content or materials, and teachers can use them only if it makes sense for them to do so. So, rather than operating as assessments *for* learning, they may be used as interim or benchmark assessments, which can be thought of as summative assessments, also known as assessments *of* learning, that can be given more than once a school year. In recent years, educational publishers have embedded assessment activities and formative feedback in digital learning systems.

Before I make recommendations about what I see as our responsibilities regarding standards for classroom assessment practice, I would like to comment on one point. I disagree with the observation that "classroom teachers would benefit from reading the *Standards*" (Plake & Wise, 2014, p. 6) because the standards are written in our technical language, not theirs. I do endorse the joint committee's statement that "promoting assessment literacy for teachers was another important goal" (Plake & Wise, 2014, p. 6). A translation of relevant *Standards* into standards for classroom assessment practice could be valuable. A good start at translating measurement concepts for classroom teachers is illustrated in a special issue of *Educational Measurement: Issues and Practice* (Benson, 2003), though some of the concepts and language in that issue are not tuned for teachers and the practical realities of classrooms. In addition, several of our colleagues have written textbooks specifically for teachers on conducting classroom assessment as assessment for learning.

So the *Standards* are not the right place and do not use the right concepts and language for teachers who need to develop classroom assessment literacy. The educational testing community should, in my view, promote the recommended practices that already exist. They appear in a number of excellent textbooks, including (with abject apologies to colleagues whose work I may have overlooked), for example, Brookhart and Nitko (2008), Chappuis, Stiggins, Chappuis, and Arter (2012), Heritage and Stigler (2010), McMillan (2014), Oosterhof (2008), Popham (2013), Russell and Airasian (2011), Taylor and Nolen (2008), and Wiliam (2011).[1]

Several professional communities and their professional associations should, in my view, promote these recommended practices and provide much needed professional development: teacher education programs, schools and school systems that contract for and provide teacher professional development, and professionals like myself who work for educational testing and publishing companies. I'm calling on teacher education organizations like the American Association of Colleges for Teacher Education, teacher associations like the American Federation of Teachers and National Education Association, NCME, and the Association of Test Publishers, as well as legislators and policy makers and advocates (e.g., the National School Boards Association) to promote these practices and make them requirements for teacher certification and professional practice. More specifically, I'm calling on the NCME Board's Outreach Committee to reach out to these organizations and get things started.[2,3]

Test Security

The agencies that sponsor and require tests—for example, the U.S. and state departments of education—have a vested interest in ensuring the integrity of the test data that they require. Test security serves to protect that integrity. Violations of test security are quite familiar to us, so temptation to achieve high test scores by fraudulent means must always have been great. The pressure to improve test scores is even higher today, especially as accountability shifts from the school level to teachers.

The *Standards*' treatment of test security is less than adequate. The *Standards* focus on protecting secure test materials, training and documentation to enable orderly and appropriate test administrations, and protecting against cheating in and outside of test administrations. They scatter discussion and standards on test security across chapters 3, 4, 6, 7, 8, 9, 10, and 12. By my count, there are 14 total standards on test security, several of them overlapping rather than unique contributions, plus commentary in chapter 4 (Test Design and Development), chapter 6 (Test Administration, Scoring, Reporting, and Interpretation), chapter 8 (The Rights and Responsibilities of Test Takers), and chapter 12 (Educational Testing and Assessment). This scattering neither provides a coherent picture on test security nor concentrated guidance to the testing communities that rely on the *Standards*. The next revision of the *Standards* should consolidate standards and commentary in a single location. Of course, that could be 15 years from now. And, whereas the standards that are present identify our roles and responsibilities regarding protecting secure test materials and preventing and looking for possible cheating, the *Standards* do not guide us on what to do if breaches of test security may have occurred.

The *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practices, 2004) includes eight statements that are directly relevant to test security and the three focuses in the *Standards*. Statement 6 in section D of the Code, Informing Test Takers, simply exhorts us to "describe procedures for investigating and resolving circumstances that might result in canceling or withholding scores, such as failure to adhere to specified testing procedures" (p. 10).

Little guidance on investigating and resolving security violations exists elsewhere. The Council of Chief School Officers' *TILSA Test Security Guidebook* (Olson & Fremer, 2013) provides some wise guidance on investigations (i.e., be conservative, respect privacy, maintain records) but none on how to conduct investigations. A chapter in the *Handbook of Test Security*, "Conducting Investigations of Misconduct" (Harris & Watkins Schoenig, 2013) provides comprehensive descriptions of types and sources of evidence (e.g., "Allowing for Explanatory Evidence from the Suspect"; see pp. 205–213) but does not address professional techniques to conduct investigations and, for example, minimize and detect evasion and lying. A recent Government Accountability Office briefing report to Education Secretary Arne Duncan on 51 state testing programs (2013) indicates that test security policies and procedures vary widely (e.g., 22 states reported that they have implemented "leading practices" [p. 3] in test security training, while four states reported none of these practices). And that's in light of other findings, including that 40 states reported allegations of cheating during 2010–2011 and 2011–2012; 33 confirmed cheating incidents; and 32 reported canceling, invalidating, or nullifying test scores because of test security violations. Many states reported feeling vulnerable to security breaches in some parts of the testing process, and 35 states reported that additional collaborations among states on test security would be useful. The U.S. Department of Education (2013) proposed some guidance on investigating alleged and actual testing irregularities, including establishing "processes that are conducive to conducting an investigation if one is required" (p. 11) and suggested that "trained personnel usually should conduct the investigation rather than school personnel" (p. 13). But where would those processes and personnel come from?

The TILSA guidebook is silent on how to resolve matters when a breach is confirmed through investigation. In another recent study, the Office of the Inspector General (2014) examined test security policies in five states. According to that report, all five states have "systems of internal control designed to prevent and detect inaccurate, unreliable, or incomplete statewide test results. However, these systems *did not always require corrective action* if indicators of inaccurate, unreliable, or incomplete statewide test results were found" (p. 1; emphasis added).

Investigation and resolution are not simple matters. Investigation is a profession with standards of practice and supporting theory and research. Higher education degree programs in criminal justice require courses in investigative techniques, and the knowledge, skills, and other abilities required of investigators are antithetical to those required of educators. For example, school administrators are trained to "collaborate with teachers" and "counsel and provide

guidance to students"[4] whereas professional investigators are trained to "prepare evidence," "interview witnesses or suspects and take statements," and "conduct in-depth investigations"[5]—two very different sets of knowledge, skills, and professional proclivities. Resolving matters when there has been a security violation, such as canceling test scores, suspending or firing staff, and revoking teaching certificates, are highly sensitive matters that may be subject to legal challenge.

Several communities could take the lead on developing comprehensive standards and a cadre of professionals to guide and support testing programs, their sponsors and authorizing bodies, and their contractors on prevention, detection, investigation, and resolution (PDIR; see Ferrara, 2014). The *Standards* and their promulgating organizations, AERA, APA, and NCME, cannot get this done alone. Testing contractors support state assessment directors with prevention and detection but, with the exception of the specialty company, Caveon (see http://www.caveon.com/services/investigations/), shy away from politically fraught investigations. State and local school systems typically are not equipped to investigate allegations of security violations. They need the assistance of professionally trained investigators when evidence suggests that a test security violation may have occurred (Ferrara, 2014). In addition, the Inspector General's report (2014) suggests that states may not take corrective action when a security violation claim is substantiated; perhaps they need guidance on when and how to take action. The educational testing community, which includes sponsors, testing program staff, and contractors, may be doing a reasonable job on prevention and detection. Not so, when it comes to investigation and resolution.

I propose that the U.S. Department of Education commission and fund an independent task force from the broader testing community (i.e., in educational, psychological, credentialing, and workplace testing) to develop coherent and comprehensive standards for test security; recommended practices for prevention, detection, and investigation; and training and certification for a cadre of experts who can be hired to conduct effective and defensible investigations for testing programs. Further, testing program sponsoring agencies and authorities should develop and implement policies to provide rigorous and fair enforcement of test security standards and strong sanctions for unprofessional behavior and unethical behavior by examinees. Perhaps the National Association of State Boards of Education can provide advice on resolving alleged security violations when results from investigations warrant action. With these recommendations, responsibility for prevention practices would be in the hands of testing programs and their contractors, statistical analysis to detect possible violations would be in the hands of psychometricians and statisticians, investigation practices would be in the hands of trained professionals, and resolution policies and procedures would be in the hands of test sponsors and authorities.

Cooperation among the educational testing community, government agencies, and related entities is required. If the U.S. Department of Education doesn't act, NCME or the National Research Council's Board on Testing and Assessment[6] should consider forming a task force or study group and seek funding to develop test security standards of practice. Who will take the first step?

Acknowledgments

My thanks to Greg Cizek, Susan Trent, Jon Twing, and Denny Way for their wise advice on the views in this commentary and to Derek Briggs. And my apologies to the joint committee for not recognizing and raising the issues in this commentary during the *Standards* revision process.

Notes

[1]And this list could be expanded by adding books on formative assessment in English language arts, mathematics, physical education, and the arts, which are easy to find via web searches.

[2]In case you're wondering, I notified the NCME Board and the Outreach Committee about my plans to call on the Outreach Committee in this commentary.

[3]I'm refraining from calling on textbook publishers to promote these recommended practices because of my and their conflict of interest.

[4]From O*Net job category 11–9032.00 – *Education Administrators, Elementary and Secondary School*; from http://www.onetonline.org/link/summary/11--9032.00

[5]From O*Net job category 13–2099.04 – *Fraud Examiners, Investigators and Analysts*; from http://www.onetonline.org/link/summary/13--2099.04

[6]In case you're wondering, I notified the staff of the Board on Testing and Assessment about my plans to call on them in this commentary.

References

Benson, J. (Ed.). (2003). Special issue on classroom assessment. *Educational Measurement: Issues and Practice*, *22*(4).

Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Upper Saddle River, NJ: Pearson.

Chappuis, J., Stiggins, R., Chappuis, S., & Arter, J. (2012). *Classroom assessment for student learning: Doing it right—Using it well* (2nd ed.). Upper Saddle River, NJ: Pearson.

Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 3–17). New York, NY: Taylor and Francis.

Ferrara, S. (2014, October). *A framework for policies and practices to improve test security programs: PDIR*. Presentation at the Conference on Test Security, Iowa City, IA.

Government Accountability Office. (2013). K-12 education: States' test security policies and procedures varied. *GAO-13–495R Test Security Leading Practices*. Retrieved October 31, 2014, from http://www.gao.gov/assets/660/654721.pdf

Harris, D. J., & Watkins Schoenig, R. R. (2013). Conducting investigations of misconduct. In J. A. Wollack & J. J. Fremer (Eds.), *Handbook of test security* (pp. 201–220). New York, NY: Routledge.

Heritage, M., & Stigler, J. W. (2010). *Formative assessment: Making it happen in the classroom*. Thousand Oaks, CA: Corwin.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: American Psychological Association. Retrieved October 31, 2014, from http://www.apa.org/science/programs/testing/fair-testing.pdf

McMillan, J. H. (2014). *Classroom assessment: Principles and practice for effective standards-based instruction* (6th ed.). Upper Saddle River, NJ: Pearson.

Office of the Inspector General. (2014). *Final audit report*. Final Audit Report ED-OIG/A07M0001. Retrieved October 31, 2014, from http://www2.ed.gov/about/offices/list/oig/auditreports/fy2014/a07m0001.pdf

Olson, J., & Fremer, J. (2013). *TILSA test security guidebook: Preventing, detecting, and investigating test security irregularities*. Washington, DC: Council of Chief State School Officers.

Oosterhof, A. (2008). *Developing and using classroom assessments* (4th ed.). Upper Saddle River, NJ: Pearson

Plake, B. S., & Wise, L. L. (2014). What is the role and importance of the revised AERA, APA, NCME *Standards for Educational and Psychological Testing? Educational Measurement: Issues and Practice*, *33*, 4–12.

Popham, W. J. (2013). *Classroom assessment: What teachers need to know* (7th ed.). Upper Saddle River, NJ: Pearson.

Russell, M. K., & Airasian, P. W. (2011). *Classroom assessment: Concepts and applications* (7th ed.). New York, NY: McGraw-Hill.

Stiggins, R. J. (2005). From formative assessment to assessment FOR learning: A path to success in standards-based schools. *Phi Delta Kappan*, *83*, 758–765.

Taylor, C. S., & Nolen, S. B. (2008). *Classroom assessment: Supporting teaching and learning in real classrooms* (2nd ed.). Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.

U.S. Department of Education. (2013). *Testing integrity symposium: Issues and recommendations for best practices*. Retrieved October 31, 2014, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2013-454

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

