Making a Difference: 2005 Evaluations and Assessment Conference. 30 November-1 December, Sydney.

An Example of how Meeting Standards Biases Institutions Toward Traditional Assessment Methods: Pre-Service Science Teacher Education

Fernando F. Padró & Marlene M. Hurley
Monmouth University, West Long Branch, New Jersey, USA

This paper touches on how pressures to meet and maintain accreditation/recognition at institutional and program levels challenge the willingness of universities to be innovative in the demonstration of student learning. Taking the example from pre-service science teacher preparation, the authors examine how institutional accreditation criteria are balanced with program-based accreditation requirements and other existing professional standards. There is a discussion on "teacher work samples," how the National Science Education Standards illustrate the need for performance-based assessment, and perform an internet research study on teacher work samples that predict the continuance of traditional testing in science courses.

Introduction

Learner outcomes and their assessment have become an important element in the accrediting process of universities and their programs in the United States. One of the underlying rationales is the belief that learner performance in the classroom is linked to institutional quality (Padró, 2005). In a sense, learner outcomes demonstrate a preference for accountability-based assessments rather than the traditional accreditation focus of quality assurance (Ewell, 2001). Both regional and programmatic level accreditation agencies and organizations set guidelines, or more prescriptive specific standards, by which programs and universities as a whole are measured for appropriateness and quality. Consequently, as Maki (2002) observes, these standards are viewed as externally mandated; as such, institutional interest ebbs and flows in relation to the timing of accreditation visits. More importantly, it seems that in order to meet the regulatory compliance element of accreditation, particularly at the programmatic level, institutions tend to focus on conservative assessment methodology. The effect is similar to the one noted by the American Productivity and Quality Center [APQC] (2004, p. 5):

> Too often, whether due to a lack of time and information, organizations find themselves picking measures that fit into a generic mold as opposed to identifying essential drivers of performance.

Walvoord (2003) suggests that the traditional modes of productivity for student learning may not be enough. The Center for Psychology in Schools and Education (1997) recommends that effective learning should include multiple assessments for diagnostic, process, and outcome purposes. Yet, it seems that universities err on the side of caution and focus more on traditional testing methodology such as traditional examinations to document student learning

and meet accreditation criteria, even in disciplines/professions that recognize the need for different assessment techniques. Performance assessment instruments such as open-ended or extended response exercises, field-based projects, or portfolios are not used in measuring learner performance. This can be due to the costs in time and planning on the part of instructors and learners (US Department of Education, Office of Research, 1993), or to a reticence on the part of universities to accept the emerging realities of competency-based approaches to assessment (Voorhees, 2001).

Brown (2004) wonders whether a university quality audit process, such as AUQA's, can become like the ritual of ISO 9000 audits in business environments. The same concerns apply to accreditation processes in the USA. The purpose of this paper is to provide a discussion of how accountability pressures seem to exert a negative influence on metrics universities use to demonstrate results in compliance to standards and other expectations. The authors use the example of current practice in the USA to illustrate how one particular strategy developed to meet identified standards seems to elicit traditional assessment mechanisms rather than encourage innovative assessments that better reflect student learning (which in itself is a contradiction of a second, supporting set of professional standards).

Learner Outcomes and Institutional Accreditation

Table 1 below provides college- and university-based accreditation criteria language specific to learner outcomes. The Middle States Commission on Higher Education (1996, p. 7) says, "[t]he ultimate goal of outcomes assessment is to examine and assess an institutions effectiveness." The process needs to be systematic. Middle States Commission on Higher Education (2002) believes measures chosen should be part of an assessment plan that provides students, faculty and others specific information about how and how much students learn. The assessment plan should begin with what universities already have in place and then generate new datasets based on multiple approaches that can be quantitative and/or qualitative in nature. The important part of this plan is that the assessments of learner outcomes should be available to planners and appropriate decision-makers. The Higher Learning Commission (2003) challenges institutions to create a culture of assessment with its accreditation criteria. Student learning for them provides a framework for evaluating learning experiences and their impact on students. The expectation is that the results "should have such credibility that they shape budgeting and planning priorities" (Higher Learning Commission, 2003, p. 3.2-9). Table 1 below provides the learner outcome criteria currently followed by these two regional accrediting agencies.

Learner Outcomes and Program Accreditation

For program-based accreditation, the emphasis still seems to be on learning goals rather than specific learner outcomes. For example, the American Assembly of Collegiate Schools of Business (2005) wants universities to set targets for their business programs that assure that learning progresses in specified directions that fit "with the goals of the degree programs, pedagogies in use, and the schools' circumstances." (American Assembly of Collegiate Schools of Business [AACSB], 2005, p. 62) Stand-alone tests or a special assessment (e.g., senior thesis) are sufficient to demonstrate a specific performance. In the area of pre-service teacher preparation, the National Council for Accreditation of Teacher Education [NCATE] (2002), the assessment system required under its Standard 2 has to reflect state professional standards as well as the school's or college's conceptual framework. Decisions about candidate performance are based on multiple assessments throughout various stages of the learner's experience ranging from admissions to transition points and program completion. Assessments in this case are used as predictors of learner (teacher candidate) success in the field.

Table 1. Identified accreditation criteria for expected evidence of learner outcome assessment data at the regional accreditation level

Middle States Commission on Higher Education	**The Higher Learning Commission**
Fundamental elements in assessing student learning: • Articulated expectations of student learning at various levels (institution, degree, program, course) consonant with institution's mission, the standards of higher education and relevant disciplines. • Plan describing student learning assessment activities undertaken by the institution, including the specific methods used to validate articulated student learning goals/objectives. • Evidence student learning assessment information is used to improve teaching and learning. • Documented use of student learning assessment information as part of institutional assessment. Optional analysis and evidence: • Assessments derived from the institution's mission and which might incorporate outcomes like cumulative learning, analytical and information skills, specific competencies, knowledge and cognitive abilities, student attitude development and growth, life skills, student activity involvement, and physical skills and techniques. • Analysis of assessment results including (where applicable, basic skills development, subject area knowledge, development of general education and lifelong learning skills), attitudes and values relating to institutional mission and programs. • Analysis of direct and indirect indicators of student achievement like persistence, graduation rates, student satisfaction, and other evidence of student goal attainment, licensure examination results, alumni satisfaction and achievement, including parity of outcomes across different student groups.	Examples of evidence: • Differentiated learning goals for undergraduate, graduate, and post-graduate programs identify the expected learning outcomes for each. • Assessment of student learning provides evidence at multiple levels: course, program, institutional. • Assessment of student learning includes multiple direct and indirect measures. • Results are available to appropriate constituencies, including students themselves. • Integration of student learning assessments processes and their data for purposes of external accountability. • Faculty members involved in defining expected student learning outcomes and creating strategies to determine the achievement of those outcomes. • Faculty and administrators routinely review university's student learning program effectiveness.

The Example of Teacher Work Samples (TWS) in Science Education as a Means of Pursuing NCATE Accreditation

To demonstrate how accreditation pressures impact institutional assessment strategies, this section discusses how "teacher work samples" (TWS) used in the preparation of science teachers encourage traditional examination results rather than fostering multiple assessment techniques set forth by the National Science Education Standards established by the National Research Council in 1996. NCATE's performance-based criteria for pre-service teacher preparation have as their goal for universities provide evidence of the teaching success of graduates and to refine university teaching techniques based upon the graduate's success.

Many education programs are using TWS's to expedite the required gathering of data for attaining and maintaining accreditation.

The National Science Education Standards

The creation of the National Science Education Standards (NRC, 1996) marked a national endeavor in the USA to standardize all teaching of science from the kindergarten through 12th (K-12) grade levels in an attempt to increase the levels of science literacy and science inquiry actively performed in schools. Science literacy implies the attainment of abilities to read, write, and discuss science with an understanding that can be applied across the natural world and to real world contextual questions, problems, and issues (e.g., Hurley, 1998). Science inquiry includes archival investigations, experimentation, and other active investigations and is assessed based upon the student's demonstrated performance:

> The inquirer must be able to pose questions and hypotheses amenable to scientific investigation and experimentation, to criticize plans for scientific investigations and experiments, to write a report about an inquiry, and to identify reliable sources of scientific information. These abilities, in turn, require information and the mental capabilities to process that information: that is, to reason scientifically (Champagne, Kouba, & Hurley, 2000, p. 454).

The NSES standards contain standards for teaching, professional teacher development, science content, programs, and systems, as well as assessment. The assessment standards moved the emphasis away from traditional testing toward multifaceted assessments:

> If the principles in the assessment standards are followed, the information resulting from new modes of assessment applied locally can have common meaning and value in terms of the national standards, despite the use of different assessment procedures and instruments in different locales. This contrasts with the traditional view of educational measurement that allows for comparisons only when they are based on parallel forms of the same test (NRC, 1996, p. 78).

Performance-based assessments are especially important components of science education at the K-12 levels due to the diversity of students who traditionally have difficulties with science and its assessment due to gender and/or ethnicity (e.g., National Science Foundation [NSF], 1996), and poor and/or lack of science instruction (e.g., Decker, 2002). Change is imperative because students from the USA performed below the international mean in science and statistically significantly lower than the international mean in problem-solving on the 2003 PISA (Programme in International Student Assessment) test scores (Organisation for Economic Co-operation and Development, 2004).

Ironically, national school reform is working to increase the challenge of using multiple assessments in science (and perhaps with other subjects, as well). The passage of the No Child Left Behind (NCLB) Act of 2001 requires high stakes testing in all subject areas at three grade levels (3-5, 6-9, 10-12) by the 2007-2008 school year (Decker, 2002). The format of these tests is left up to the states, but they have largely been using multiple-choice tests rather than the "rich assessments called for in the NSES" (Decker, 2002, p. 50).

Using TWS's to Pursue NCATE Accreditation

The preference shown previously for traditional examinations in the name of ensuring that accrediting agencies feel comfortable with the documentation of learner outcomes is

highlighted when looking at a process called "teacher work sample" (TWS) for programs pursuing NCATE accreditation or reaffirmation. Beginning in 2001, the Washington, DC-based National Council for Accreditation of Teacher Education (NCATE) phased in "performance-based" standards that were fully in place by 2005 for all 600 NCATE-approved teacher education schools in the United States. The goal of these standards is for education schools to provide evidence of the teaching success of their graduates and to refine their techniques based upon their graduate's success.

The teacher work sample, developed by Schalock at Western Oregon University about 25 years ago, is an extremely structured unit of curriculum written and taught by teacher candidates as a field practicum component of a degree program leading to a teaching certificate. A TWS provides data that reflects both the professional capabilities of the teacher candidate and the learning of students instructed by the teacher candidate. The TWS uses narrative explanations of the seven different parts of the unit (context, learning goals, assessment plan, design for instruction, instructional decision-making, analysis of student learning, and reflection with self-evaluation) to promote teacher candidate reflection and the use of sound pedagogies. Furthermore, the teacher work sample is thought to expedite the required gathering of data for attaining and maintaining accreditation (Olson, 2005).

TWS's are controversial because of the time demands on the teacher candidate, the cooperating teacher, and the supervising teacher. The priority seems to be on completing the TWS and meeting its expectation rather than on meeting classroom responsibilities fully (Olson, 2005). TWS's are touted as a performance-based assessment of the teacher candidate (McConney, Schalock, & Schalock, 1998); yet, they seem encourage candidates to use traditional and low level assessments of their students with their pre- and post-test reporting requirements. Efforts to use the suggested gain scores as measures of the learning gains of students had a "negative impact on the significance of the learning goals and the quality and types of assessments candidates employed in their work samples" (Denner, Salzman, & Bangert, 2001, p. 304).

An Internet Study of TWS

The authors found at least 100 websites were found addressing some aspect of TWS. These were each searched for two types of information: (1) instructions for writing the "assessment plan" section of the TWS and (2) examples of TWS Assessment Plans for teaching science. Twenty-two web sites from various institutions contained enough instructions for writing the Assessment Plan portion of the unit (specifically the major pre- post assessment piece) for the authors to code the instructions with "T" for traditional (pre-/post-tests), "N" for noncommittal (traditional and non-traditional assessments), and "P" for performance-based assessment. Out of the 22 web sites, 10 were classified as "T" for traditional sets of instructions, 10 as "N" for recommending noncommittal sets of instructions, and 2 proposed performance-based sets of instructions.

The authors found a total of 10 examples of TWS's from various universities that were written and graded as "acceptable" for teaching science units. Table 2 provides the results of the analysis of the 10 teacher work samples. Note that only two out of the ten TWS's utilized multiple assessment types, although one of these did include a traditional testing scheme. Seven out of the ten TWS's utilized traditional multiple choice tests as the basis of their pre- and post-test assessment structure.

Table 2. Ten TWS science assessments showing type of assessments prepared and taught by teacher candidates during practicals.

TWS Grade Level	Pre- and Post- Assessments	Assessment Type
Elementary, level 3	Multiple choice, matching, true-false	Traditional paper & pencil
Kindergarten	Oral recall of knowledge	Traditional oral
* Elementary, level 2	Oral, written, & drawings	Traditional paper & pencil with oral problem-solving
Elementary, level 3	Multiple choice, fill-in, true-false	Traditional paper & pencil
Elementary, level 2	Multiple choice	Traditional paper & pencil
Elementary, level 1	Multiple choice	Traditional paper & pencil
Elementary, level 4	Multiple choice, true-false	Traditional paper & pencil
Secondary, level 9	Multiple choice, 3 knowledge essays	Traditional paper & pencil
* Kindergarten	Oral	Performance-based with rubric
Secondary	Multiple choice, matching	Traditional paper & pencil

* Two (20%) of ten teacher work samples examined used nontraditional or performance-based assessments.

As previously stated, TWS uses narrative explanations of the seven different parts of the unit to promote teacher candidate reflection and use of sound pedagogies; however, the TWS structure suggests page limits be applied to each of the seven areas. The message is that they are to reflect, but not too deeply, with a maximum of 20-25 narrative pages for all seven parts seeming to be the suggested norm. Consequently, the products that are derived from the TWS materials are superficial and poorly written. The focus is to either "squeeze" the narrative into the page allocations or eliminate narrative in lieu of meeting the page requirements. TWS sample units studies that yield more complete data about learning outcomes are these where the teacher candidates ignored the institution's page limitations and provided more complete research, reporting, analyses, and evaluation.

Concluding Remarks

In the USA, the role of external reviewers of higher education institutions is to *clarify* what university processes are doing and *verify* that the processes lead to reported results that meet accreditation criteria and related performance standards. Reviewers need to recognize that learning is taking place that is relevant with professional and vocational requirements, personal and social policy needs. While they look for compliance, examiners are not charged with stifling innovation, or being prescriptive about programming. What they do not want to see are processes and measures that are generic in nature, as APQC (2004) warns.

Twigg (2005) recommends that continuous assessment/feedback and mastery learning become one aspect of redesigning courses to promote and ensure quality. It not only promotes student responsibility in learning but educational practice will likely improve when the learner becomes the primary focus (Center for Psychology in Schools and Education, 1997). Yet, the example of the TWS in the area of science education exemplifies how institutions tend to gravitate toward generic or traditional assessments mechanisms to make sure that they are able to demonstrate to external reviewers that their processes comply with externally derived or imposed criteria or standards (Table 2).

This discussion has centered on "teacher work samples" in science education because what is found here is a combination of differing programmatic goals needing to be met. There is an NCATE Standard for program assessment and accountability that has to be met in order to achieve or maintain pre-service teacher program accreditation. There are discipline standards, the National Science Education Standards, that want to influence the preparation of future primary and secondary level science teachers. And there is an overt or embedded requirement

at the institutional level to demonstrate student learning through the use of learner outcomes. The question that universities with pre-service teacher preparation programs is not "are these three expectations compatible?" but "how can we do all three or at least ensure that accreditation/approval is maintained?" The compliance perspective pressures institutions to use methodology that has stood the test of time and will pass muster, at least at a *de minimis* level.

It seems TWS's support the "old" instructor-centered environment rather than the current view of the learner-centered environment (Padró, 2005) as exemplified by the preponderance of these traditional assessment types. Writing a multiple-choice test or using the textbook publisher's multiple choice tests was for many years the non-thinking teacher's favorite assessment. There is concern in the literature that TWS's are perhaps unintentionally halting or reversing the progress that has been made in science assessment (e.g., Airasian, 1997; Darling-Hammond, 1998).

This paper is by no means an exhaustive review of the literature on TWS's, learner outcomes, or accreditation either within the USA or elsewhere. Instead, this is an illustration on the pressures that institutions have to ensure accreditation or other forms of recognition as well as compliance anywhere in this world. Innovation can be risky although desired. Universities, for a number of other reasons not discussed here, often feel that they cannot take risk that could challenge the maintenance of their accreditation because they do not have the time, resources, interest, or a clear understanding of what is required and expected of them.

Notes on Contributors

Fernando Padró is Associate Professor of Education Leadership at Monmouth University specializing in higher education administration. Research interests are in institutional quality assurance at universities and the role of accreditation. A former Baldrige National Quality Award Examiner, he is a Project AQIP reviewer for The Higher Learning Commission.

Marlene Hurley is Assistant Professor in Science Education at Monmouth University. She holds a Ph.D. in Science Education from SUNY-Albany, with more than 20 years in science education at the middle school, secondary, and higher education levels. She is also a science education consultant and program evaluator. Her research interests include plant science, animal science, environmental science, the integration of science with other subjects, and problem-based learning.

Addresses for Correspondence. Dr. Fernando F. Padró, Monmouth University, McAllan Hall 105, West Long Branch, N.J. 07764, USA. E-mail: fpadro@monmouth.edu

References

Airasian, P. (1997). *Classroom assessment* (3rd ed.). New York: Macmillan.

American Assembly of Collegiate Schools of Business (2005). *Eligibility procedures and accreditation standards for business accreditation*. Tampa, FL: Author. Retrieved July 13, 2005 from http://www.aacsb.edu/accreditation/standards.asp

Brown, A. (2004). Directions in quality management in Higher Education in Australia. *7th Toulon-Verona Conference Proceedings: Quality in Services in Higher Education; Health Care; Local Government, 2-3 September 2004*, 119-123. Toulon, FR: Université du Sud Toulon-Var.

Center for Psychology in Schools and Education. (1997*). Learner-centered psychological principles: A framework for School Redesign and Reform.* Washington, D.C.: American Psychological Association. Retrieved July 10, 2004 from http://www.apa.org/ed/lcp.html

Champagne, A. B., Kouba, V. L., & Hurley, M. (2000). Assessing inquiry. In J. Minstrell & E. H. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 447-470). Washington, DC: AAAS.

Darling-Hammond, L. (1998). Standards for assessing teaching effectiveness are key. *Phi Delta Kappan, 79*(6), 471-472.

Decker, M. (2002). Science in the age of accountability. *Journal of Education, 183*(3), 47-52.

Denner, P. R., Salzman, S. A., Bangert, A. W. (2001). Linking teacher assessment to student performance: A benchmarking, generalizability, and validity study of the use of teacher work samples. *Journal of Personnel Evaluation in Education, 15*(4), 287-307.

Ewell, P.T. (2001, February). *From "assessment" to competency-based quality assurance*. Paper presented at the Designing Real Educational Alliances in Missouri [DREAM] Conference.

Higher Learning Commission (2003). *Handbook of accreditation* (3rd ed.). Chicago, IL: Author. Retrieved January 15, 2005 from http://AQIP.org

Hurley, M. M. (1998). Science literacy: Lessons from the first generation. *Research Matters to the Science Teacher, 9801.* Retrieved from the NARST website: www.educ.sfu.ca/narstsite/publications/research

Maki, P. (May 2002). Moving from paperwork to pedagogy: Challenging intellectual curiosity into a commitment to assessment. *AAHE Bulletin*, *54*(9), 3-5.

McConney, A. A., Schalock, M. D., & Schalock, H. D. (1998). Focusing improvement and quality assurance: Work samples as authentic performance measures of prospective teachers' effectiveness. *Journal of Personnel Evaluation in Education, 11,* 343-363.

Middle States Commission on Higher Education (2002). *Characteristics of excellence in higher education*. Philadelphia, PA: Author. Retrieved September 15, 2004 from http://www.msache.org

Middle States Commission on Higher Education (1996). *Framework for outcomes assessment*. Philadelphia, PA: Author.

National Council for Accreditation of Teacher Education (2002). *NCATE Unit Standards (2002 Edition)*. Retrieved from January 15, 2005 from http://www.ncate.org/public/unitStandardsRubrics.asp?ch=4#stnd2

National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.

National Science Foundation (1996). *The learning curve.* Washington, DC: Author.

Olson, L. (2005). Education schools use performance standards to improve graduates. *Education Week, 24*(36), 1-4.

Organisation for Economic Co-operation and Development (2004). *First results from PISA 2003.* Paris: Author.

Padró, F.F. (2005). Learner outcomes: A component of a new paradigm of instruction, a significant measure of institutional quality assurance or an object of political control threatening faculty and institutional autonomy? *Monash University 2004 Evaluation Forum Proceedings, November 24-25, 2004, 116-126.* Melbourne: Monash University Centre for Higher Education Quality.

Twigg, C.A. (June 2005). Course redesign improves learning and reduces cost. *The National Center for Public Policy and Higher Education Policy Alert*, June 2005. Retrieved September 10, 2005 from http://www.highereducation.org/reports/pa_core

US Department of Education, Office of Research (1993). Performance Assessment. In *Education Consumer Guide* (No. 2). Washington, DC: Author. Retrieved July 10, 2005 from http://www.ed.gov/pubs/OR/ConsumerGuides/perfasse.html

Voorhees, R.A. (2001). Competency-based learning models: A necessary future. In R.A Voorhees (Ed.), *Measuring what matters: Competency-based learning models in higher education: New directions for institutional research, 110*, 5-13. San Francisco: Jossey-Bass.

Walvoord, B.E. (2003). New modes of productivity for student learning. In J.E. King, E.L. Anderson, & M. E. Corrigan (Eds.). *Changing student attendance patterns: Challenges for policy and practice: New Directions for Higher Education, 121*, 35- 49. San Francisco: Jossey-Bass.