

A Call for Improved Teacher Assessment Training

Marie A. Revak
Florida Institute of Technology

ABSTRACT. This literature review reports on the status of teacher assessment training and assessment practices in relation to the recommendations set forth by the National Council of Teachers of Mathematics (NCTM) Curriculum and Evaluation Standards for School Mathematics (1989). Present methods of assessment used by teachers are not in congruence with the NCTM standards. Furthermore, teachers have not been adequately trained to design, administer, score, or interpret the alternative assessments recommended by the NCTM. Recommendations are made for improved teacher assessment training and research to determine the effectiveness and usefulness of new assessment training programs.

Over the past several decades, research on classroom assessment has typically been confined to the study of standardized tests, student responses to tests, and the interpretation of test scores by teachers, school administrators, and the general public (Haladyna, Nolen, & Haas, 1991; Pedulla, Airasian, & Madaus, 1980; Rudman, 1987; Stiggins, 1991b). More recently, the topic of assessment reform has found a place in the literature (McLean, 1990, Romberg & Wilson, 1992; Shavelson, Baxter, & Pine, 1991; Stiggins, 1988), along with studies investigating the use of assessments created in the name of reform (Borasi & Rose, 1989; Dellinger, 1992; Miller, 1992; Stahle & Mitchell, 1993, Usnick & Brown, 1992). Although the limitations of standardized tests and the benefits of new, alternative forms of assessment are well-documented, current research has not addressed the link connecting teachers to the assessment reform movement. This link is teacher assessment training.

Because training is a key component of successful reform, this paper examines the new assessment standards put forth by the National Council of Teachers of Mathematics (NCTM) as they relate to the present state of teacher assessment training and teachers' current use of the assessments advocated by the NCTM. This paper concludes with implications for research and educational practice and recommendations concerning assessment training for pre-service and inservice teachers.

Mathematics Assessment Reform

In 1989, the NCTM issued the Curriculum and Evaluation Standards for School Mathematics as a basis for mathematics education reform. This key document takes its cues from constructivist theory, in which new learning is constructed from what is already known or believed. To quote Etchberger and Shaw (1992), "making sense of information and experience is the construction of knowledge" (p. 411).

In addition to setting standards for mathematics instruction in several key content areas, the NCTM Standards specify three canons for student evaluation: (1) alignment of the assessment methods and tasks with the goals, content, and emphases of the curriculum; (2) multiple sources of assessment information, including tasks that demand a variety of mathematical thinking set in a variety of contexts; and (3) the selection of appropriate assessment methods depending on the type of information required, the intended use of the information, and the developmental level of the student. In discussing multiple assessment techniques, the NCTM Standards recommend the use of written, oral, and demonstration formats to authentically assess desired outcomes and improve the quality of decisions concerning students' learning.

Publication of the NCTM Standards does not guarantee their implementation. Obstacles to be overcome include the prevailing tests, current classroom testing practices, and teachers' ability to create and use the types of assessments recommended by the Standards.

The NCTM contends that existing tests cannot measure the student outcomes identified in the Standards. "The majority of the educational outcomes we value for students cannot be translated into objective paper and pencil tests" (Stiggins, 1991b). In addition, current tests fail to reflect the goals of the curriculum. The "Standards for Teacher Competence in Educational Assessment of Students" warn against using assessments that lack curriculum validity (American Federation of Teachers, National Council on Measurement in Education, & National Education Association, 1990). Romberg and Wilson (1992) analyzed six widely used standardized mathematics achievement tests and found that an average of 89 percent of the items tested procedures and that the tests did not cover the range of content specified in the Standards. Romberg and Wilson (1992) also report a follow-up study of new tests being used in California, Connecticut, South Carolina, Massachusetts, Vermont, Britain, Australia, France, Korea, the Netherlands, and Norway. In reviewing these tests, they found problems that were rich, engaging, interesting, and in congruence with NCTM recommendations, but conclude that the use of such tests is still uncommon (Romberg & Wilson, 1992).

The publishing of the NCTM Standards and modest changes in standardized tests may not impact the current classroom assessment practices of teachers. Allal (1988) found that although a major reform of mathematics instruction was introduced in French-speaking Switzerland 15 years ago, there has been very little impact on teachers' evaluation practices. Stiggins and Bridgeford (1985) found evidence that teachers do not tend to vary

their assessments, even if the purpose of assessment varies. In their study, only a few of the 228 surveyed teachers planned to use a new type of assessment in the future.

As mathematics assessment moves away from decades of dependence on objective paper and pencil tests, classroom teachers will bear the lion's share of the workload. However, teachers may not be equipped to deal with these tremendous changes in student assessment. In a study of four teachers who have begun to implement the instructional practices suggested by the NCTM Standards, Edgerton (1992) found that although each of the teachers expressed a desire to make changes in their assessment practices, they felt constrained by time, lack of alternative models, lack of opportunities to share ideas with colleagues, and the expectation to cover a prescribed number of course topics.

Assessment Training

Background

Teachers report spending 10 to 30 percent of each day in the evaluation and assessment of student learning, but they have not been trained to do the job well (Schafer & Lissitz, 1987; Gullickson, 1984; Stiggins, 1988). Carter's (1984) study of secondary reading and language arts teachers concludes that many were not familiar with basic test writing principles, nor did they fully appreciate the impact of their assessment decisions. Carter (1984) reports that "most problematic for teachers were items testing higher-order skills, specifically skills of inference and prediction" (p. 59). A study by Taylor (1991) had similar results. Taylor's (1991) analysis of a sample of test items submitted by algebra and geometry teachers indicated that most teachers were testing at a knowledge or skill level but thought they were testing higher order thinking skills. Stiggins, Griswold, and Wiklund (1989) found that teachers' written and oral tests were dominated by recall questions. However, they report that those who were trained in assessment asked a greater percentage of higher order thinking questions than those not trained.

Indeed, studies highlighting the lack of adequate teacher assessment training far outweigh those devoted to documenting particular instances of such training (Allal, 1988; Carter 1984; Gullickson 1984, 1986; Schafer & Lissitz, 1987; Stiggins, 1988, 1991a, 1991b, 1993; Stiggins & Bridgeford, 1985; Stiggins & Conklin, 1992).

Educational Measurement Courses

A Northwest Regional Educational Laboratory (NWREL) study found that a majority of schools offering degrees leading to teacher certification do not require a course in educational measurement and evaluation. The study investigated teacher-training curricula at 14 major teacher training colleges and universities in six Pacific Northwest states. Only nine (combined undergraduate and graduate) courses focused on assessment, and, of these nine, only three were required for teacher certification (Stiggins & Conklin, 1992). A similar study by Schafer and Lissitz (1987) surveyed over 400 American Association of Colleges for Teacher Education member institutions and found that most

schools did not require a formal course in educational measurement for certification. The same study revealed that key measurement topics were not covered in any required course work in a large number of curricula.

Of the teachers who have taken a typical measurement and evaluation course as part of their pre-service requirement, many indicate that the course did not prepare them to confidently conduct classroom assessments (Gullickson, 1986). Gullickson (1984) surveyed third, seventh, and tenth grade teachers and concluded that "the average teacher does not perceive college courses to be pertinent to his/her classroom testing needs" (p. 245). When Stiggins and Conklin (1992) studied the educational assessment courses offered at the major teacher training colleges of the Pacific Northwest, they found that course and text content typically emphasized statistical measures of validity and reliability of standardized achievement tests and rarely included any discussion about providing feedback, setting assessment policies, or using test results. Gullickson (1986) surveyed both pre-service teachers and professors in university education departments and found that the two groups disagree strongly on the importance of three key content areas: statistics, alternative assessment activities (such as rating scales, observation, sociograms, anecdotal records, and class discussion), and formative and summative evaluation. The professors, as a rule, place a heavy emphasis on the statistics of standardized tests. The pre-service teachers, on the other hand, want more training on creating their own tests and conducting performance evaluations and observations (Gullickson, 1986).

Limited Options for Teachers

Teachers who wish to increase their knowledge and use of alternative assessments without the benefit of training have limited options. Recently, Marshall and Thompson (1994) reviewed six of many current books focusing on assessment. They found that, although all six books document what is wrong with assessment and promote the role of teachers in assessment reform, the volumes offer little in the way of practical examples for classroom use.

Stiggins (1988) reports that most medium-sized and small school districts do not employ educational measurement specialists. Teachers cite their colleagues, who are also predominantly untrained, as their key source in learning about assessment strategies (Stiggins, 1988). In reviewing research relating to teacher assessment training, no evidence of training in consonance with the NCTM Standards was found.

Multiple Means of Assessment

Background and Rationale

The NCTM argues that a new constructivist philosophy for mathematics education requires new methods for assessing student competence. The major emphases of the NCTM Standards document are problem solving, communication, and ties to real world context and other content areas (NCTM, 1989). Many methods must be employed to

assess mathematics growth and power across these domains. Wolf (1993) summarized this philosophy: "If students are confined to a narrow range of activities, the assessments carried out will be equally narrow in their portrayal of students--and teachers will have few insights into the range of talents and interests possessed by their students" (p. 521). The "Standards for Teacher Competence in Educational Assessment of Students" developed by the American Federation of Teachers, National Council of Measurement in Education, and the National Education Association (1990) follow suit in stating that teachers need to be acquainted with a broad range of assessment options and know the strengths and weaknesses of each.

In addition to using a variety of assessment options, Wolf (1993) advocates assessment in diverse contexts over longer periods of time, and McLean (1990) recommends adding some assessment of group processes to increase the authenticity of assessments.

Teachers' Use of Assessment Methods

Although elementary and secondary teachers surveyed by Gullickson (1984) agreed that students should not be evaluated exclusively by tests, about half of those teachers used tests as their primary basis for student grades. Taylor (1991) found that few of the algebra and geometry teachers surveyed used group projects, journals, or portfolios in the assessment of students. A study by Stiggins and Conklin (1992) found that although teachers use a variety of assessment techniques in their classrooms, they also report concerns about the quality of the assessments they create. In addition, Stiggins and Bridgeford (1985) found that many teachers use unsound performance assessment practices, such as (1) relying on mental note taking, (2) failing to plan and document performance assessment criteria in advance, and (3) using too few observations to make judgments on student performance.

Allal (1988) found that teachers often use assessments other than large written tests as adjustments to test averages when assigning trimester grades. Although about half of the teachers in Allal's (1988) study report using quizzes and tallies (records of various daily tasks), only 16 percent reported use of assessments such as individual and group research projects and problem solving situations. Allal (1988) found that teachers used a combination of many qualitative factors to make promotion decisions but that the teachers "have no systematic means of record keeping for these factors" (p. 48).

Adopting Alternative Assessments

Many studies report positive experiences when one alternative assessment method is adopted (Borasi & Rose, 1989; Larter, 1991; Long & Ben-Hur, 1991; Miller, 1992; Simon, 1992; Stahle, 1993; Usnick & Brown, 1992). Of the studies reviewed, none discussed how teachers were trained to implement the new assessments. Only one example was found where constructivist assessment methods were implemented in a teacher

education course (Stahle & Mitchell, 1993). No research was found that explored the impact of adopting a variety of assessment techniques.

Implications

There is a genuine need for changes in the way teachers are trained to assess students. The restructuring of university level teacher education may eventually fill the need for assessment training of future teachers. The assessment training needs of current classroom teachers must be met in other ways.

Typical teacher training may not be the solution. Duffy and Roehler (1986) studied constraints on teacher change in implementing new instructional methods and conclude that "short term innovation and contextually-isolated methods courses will probably have little effect" (p. 57). Instead, they recommend longitudinal inservice training based in real classroom contexts.

Thus, assessment training must be carefully designed to meet the needs of teachers. Gullickson (1986) states, "persuasion of teachers [to study assessment] will require that we determine which measurement tools and strategies are both practical and beneficial" (p. 354). The training must begin with an analysis of teachers' concerns and problems. It must provide the necessary knowledge, tools, and materials and be set in a practical classroom context. Allal (1988) recommends explicit training in: (1) how to design and use simple instruments (checklists, matrices, charts, and coding systems) for recording qualitative data based on informal observations, (2) how to avoid biases in evaluating students, and (3) how to develop and use techniques for combining quantitative and qualitative assessment data.

Quality assessment training for teachers can set the stage for authentic student evaluation and informed decision making. Once assessment training programs have been established, researchers should determine the usefulness and effectiveness of the training in meeting the recommendations put forth by the NCTM (1989).

References

- Allal, L. K. (1988). Quantitative and qualitative components of teachers' evaluation strategies. *Teaching and Teacher Education*, 4, 41-51.
- American Federation of Teachers, National Council on Measurement in Education, National Education Association. (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30-32.
- Borasi, R., & Rose, B. (1989). Journal writing and mathematics instruction. *Educational Studies in Mathematics*, 20, 347-365.

- Carter, K. (1984). Do teachers understand principles for writing tests? *Journal of Teacher Education*, 35, 57-60.
- Dellinger, D. D. (1992). Utilizing alternate formats for testing arithmetic story problems. *Dissertation Abstracts International*, 53, 433-A.
- Duffy, C., & Roehler, L. (1986). Constraints on teacher change. *Journal of Teacher Education*, 37, 55-59.
- Edgerton, R. T. (1992). The assessment practices of teachers who have begun to implement the instructional practices suggested in the NCTM Standards Documents. *Dissertation Abstracts International*, 53, 747-A.
- Etchberger, M. L., & Shaw, K. L. (1992). Teacher change as a progression of transitional images: A chronology of a developing constructivist teacher. *School Science and Mathematics*, 92, 411-417.
- Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests. *Journal of Educational Research*, 77, 244-248.
- Gullickson, A. R. (1986). Teacher education and teacher-perceived needs in educational measurement and evaluation. *Journal of Educational Measurement*, 23, 347-354.
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.
- Larter, S. (1991). *Benchmarks: The development of a new approach to student evaluation*. Toronto: Toronto Board of Education.
- Long, M. J., & Ben-Hur, M. (1991). Informing learning through the clinical interview. *Arithmetic Teacher*, 38(6), 44-46.
- Marshall, S. P., & Thompson, A. G. (1994). Assessment: What's new--and not so new--a review of six recent books. *Journal for Research in Mathematics Education*, 25, 209-218.
- McLean, L. D. (1990). Time to replace the classroom test with authentic measurement. *The Alberta Journal of Educational Research*, 36, 78-84.
- Miller, L. D. (1992). Teacher benefits from using impromptu writing prompts in algebra class. *Journal for Research in Mathematics Education*, 23, 329-340.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston: Author.

- Pedulla, J. J., Airasian, P. W., & Madaus, G. F. (1980). Do teacher ratings and standardized test results of students yield the same information? *American Educational Research Journal*, 17, 303-307.
- Romberg, T. A., & Wilson, L. D. (1992). Alignment of tests with the standards. *Arithmetic Teacher*, 40, 18-22.
- Rudman, H. C. (1987). Testing and teaching: Two sides of the same coin? *Studies in Educational Evaluation*, 13, 73-90.
- Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education*, 38(3), 57-63.
- Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessment in science. *Applied Measurement in Education*, 4, 347-362.
- Simon, M. K. (1992). Alternative curriculum and assessment at an alternative high school. *Mathematics Teacher*, 85, 671-674.
- Stahle, D. L., & Mitchell, J. P. (1993). Portfolio assessment in college methods courses: Practicing what we preach? *Journal of Reading*, 36, 538-542.
- Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan*, 69, 363-368.
- Stiggins, R. J. (1991a). Assessment literacy. *Phi Delta Kappan*, 72, 534-539.
- Stiggins, R. J. (1991b). Facing the challenges of a new era of educational assessment. *Applied Measurement in Education*, 4, 263-273.
- Stiggins, R. J. (1993). Two disciplines of educational assessment. *Measurement and Evaluation in Counseling and Development*, 26, 93-104.
- Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22, 271-286.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany: State University of New York Press.
- Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*, 26, 233-246.

- Taylor, L. J. C. (1992). Survey of algebra and geometry teachers' assessment practices and an investigation of publisher-provided testing materials. *Dissertation Abstracts International*, 52, 3851-A.
- Usnick, V. E., & Brown, S. (1992). Assessing elementary students' mathematical thinking through interpersonal process recall. *Focus on Learning Problems in Mathematics*, 14, 56-62.
- Wolf, K. P. (1993). From informal to informed assessment: Recognizing the role of the classroom teacher. *Journal of Reading*, 36, 518-23.