3. The Development of Standards for Teacher Competence in Educational Assessment of Students

James R. Sanders  
*Western Michigan University*

Suzanne R. Vogel  
*Western Michigan University*

Follow this and additional works at: [http://digitalcommons.unl.edu/burosteachertraining](http://digitalcommons.unl.edu/burosteachertraining)
THE DEVELOPMENT OF STANDARDS FOR TEACHER COMPETENCE IN EDUCATIONAL ASSESSMENT OF STUDENTS

James R. Sanders and Suzanne R. Vogel

Western Michigan University

There has been a long history of concern about the quality of student assessments and their use by educators, and rightly so. Test scores, grades, informal measurements, and other forms of assessment typically have been weighted heavily in decisions about students, programs, and policies. Malpractice in student assessment can have detrimental and irreversible consequences affecting human lives and school programs. Assessment is defined here as the process of obtaining information that is used to make educational decisions about students; to give feedback to students about their progress, strengths, and weaknesses; to judge instructional effectiveness and curricular adequacy; and to inform policy.

The National Council on Measurement in Education (NCME) studied the feasibility of credentialing measurement experts in education, and concluded that because the practice of measurement and assessment is so pervasive in education and takes on so many different forms, it would be much too costly to develop credentialing procedures for every type of assessment practice (Sanders, 1987).
an alternative, the NCME undertook the development of standards of assessment competence for major practitioner roles in education: classroom teachers, school administrators, counselors, testing directors, curriculum specialists, and others. In 1987 the NCME invited three other professional associations to collaborate on the development of standards for classroom teachers, the largest practitioner group and the one that uses student assessments most frequently. Similar collaborative projects, focused on other educational practitioners, are expected to follow.

The collaborators on the teacher standards were three associations directly involved in the preparation and professional development of classroom teachers: the American Association of Colleges for Teacher Education (AACTE), the American Federation of Teachers (AFT), and the National Education Association (NEA). The report of this project was published in 1990.

In the remainder of this chapter we will review selected literature on teacher preparation in student assessment: (a) how classroom teachers use measurement and student assessments in the classroom, (b) what experts have said teachers need to know about measurement and student assessment, and (c) the status of training prospective teachers in student assessment. We will then describe the standards developed by the four collaborating associations, and conclude with a brief discussion of work that still needs to be done to improve the quality of student assessments and their use in education.

LITERATURE ON TEACHER PREPARATION IN STUDENT ASSESSMENT

The need for developing standards to guide teachers’ professional preparation and in-service training in assessment was recognized as far back as 1912 (Starch & Elliot, 1912), and has been building since 1967 when Samuel Mayo presented his report, Pre-service Preparation of Teachers in Educational Measurement, and David Goslin wrote Teachers and Testing. The importance of assessment competence for teaching was highlighted by Rudman, Kelly, Wanous, Mehrens, Clark, and Porter (1980), who described the necessity for teachers to use a variety of assessment methods in order to make appropriate decisions about student grading, grouping, placement, and instruction. The ability to use information properly when making important student, instructional, or curricular decisions is an integral part of professional teaching practice. Research has consistently revealed, however, that the preparation of teachers at most universities in the area of assessment is either inadequate or totally absent (Noll, 1955; Roeder, 1972,
3. DEVELOPMENT OF STANDARDS

1973; Schafer & Lissitz, 1987, 1988). This is true, in spite of research documenting that practicing teachers spend a substantial portion of their time in activities related to student assessment (Stiggins, 1988). In addition, training in student assessment procedures has been shown to be important to teachers (Borg, Worthen, & Valcarce, 1986).

How Classroom Teachers Use Measurement and Student Assessment in the Classroom

Gullickson (1985) conducted a survey of 295 South Dakota teachers to determine the relationship, if any, between 11 student evaluation techniques, grade level, and curriculum area. His study showed that the most highly rated techniques across all grade levels and curricula were objective teacher-made tests, discussion, and papers/notebooks.

According to Gullickson’s 1985 report, elementary teachers tend to rely on several evaluation techniques of pupil progress. “Class discussion, evaluation of student papers, and evaluation of student behavior all are seen to hold a higher priority than tests” (p. 99). The elementary teachers do tend to give more credence to the results of standardized objective tests than do junior and senior high teachers.

According to the results of Gullickson’s survey, secondary (junior and senior high) teachers tend to rely on fewer evaluation techniques, with teacher-made objective tests being the method of choice. Secondary teachers reported that they use essay tests much more frequently than do elementary teachers.

In a more extensive survey of classroom teachers in South Dakota (336 respondents), Gullickson investigated purposes for testing, frequency of testing, sources of test items, and preferred methods of measurement. The findings of this study are consistent with the previous study in that generally teachers rated teacher-made objective tests most highly. Secondary teachers again placed significant emphasis on essay tests. These evaluation techniques were followed in order by standardized objective tests and oral quizzes (Gullickson, 1982).

Teachers reported using tests frequently, with 95% indicating weekly use of tests. Gullickson’s study indicated that teachers spend a great deal of time in test-related activities, with the estimated average time spent in such activities being 190 minutes per teacher-made objective test. Assuming that teacher-made objective tests are administered on a weekly basis, this translates into about one-half teacher day per week spent on test-related work.
Teachers reported that they author their own test items 93% of the time, use items from textbook publishers 60% of the time, and use other published test items 23% of the time. When asked to indicate the types of items normally used on their tests, 92% indicated short answer/completion, 77% matching, 76% multiple choice, 67% true/false, and 58% essay, with 31% of elementary teachers using essay as opposed to 69% of secondary teachers. Teachers indicated that about 75% of their course content is covered by their teacher-made objective tests.

The great majority of teachers in Gullickson’s study indicated that the following test administration conditions are the norm:

- Students may not interact.
- Students may not use resource materials.
- Students may not use calculators, except in senior high science courses where 40% of teachers allow their use.
- Tests are not speeded.

Sixty-four percent of the teachers reported that they do not use separate answer sheets.

The overwhelming majority of teachers (97%) reported that they always or usually score their own tests. Only 55% report that they always or usually provide written comments on tests. The vast majority of teachers (90%) use total score as the only means of test analysis. Forty-two percent of the teachers use score range. Mean, median, and standard deviation are used by relatively few teachers in test analysis. Roughly one third of the teachers report analysis of item difficulties and test reliability.

Teachers reported that they generally grade (95-97%) their own tests, and 94% return tests promptly to students (within 2 days). Relatively little time is spend during class time for posttest review (Gullickson, 1982).

Gullickson’s results confirm many of the findings of Fleming and Chambers (1983), who conducted systematic analyses of teacher-made tests in the Cleveland, Ohio Public Schools in response to a federal court order for desegregation. The authors made the following observations about how Cleveland teachers test:

First, teachers use short-answer questions most frequently in their test making. Second, teachers, even English teachers, generally avoid essay questions, which represent slightly more than one percent of all test items reviewed. Third, teachers use more matching items than multiple-choice or true-false items. Fourth, teachers devise more test questions to sample knowledge of facts than of any other behavioral categories studied. Fifth, when categories related
to knowledge of terms, knowledge of facts, and knowledge of rules and principles are combined, almost 80% of the test questions reviewed focus on these areas. Sixth, teachers develop few questions to test behaviors that can be classified as ability to make applications. Seventh, comparison across school levels shows that junior high school teachers use more questions to tap knowledge of terms, knowledge of facts, and knowledge of rules and principles than elementary or senior high school teachers do. Almost 94% of their questions address knowledge categories, contrasted with 69% of the elementary school teachers’ questions. Finally, at all grade levels, teacher-made mathematics and science tests reflect a diversity of behavioral categories, since they typically feature questions in all six behavioral categories. (p. 32)

Fleming and Chambers (1993) found that teachers generally used one-page tests that were usually neat in appearance, but because of poor quality reproduction were sometimes difficult to read. Teacher-made tests often did not contain clear directions, and were found to have errors in punctuation and spelling nearly 20% of the time. Teacher-made tests lacked indication of point values for test items in most cases, which suggests to the authors that “teachers may not be visualizing their tests as a means for quantifying students’ performance as a measure of students’ learning” (p. 36).

The Cleveland study indicated some problems with item construction. For example, multiple-choice item stems might be only one or two words; short answer/completion items might be unclear; multiple-choice items might have more than one defensible correct response. The authors concluded that their review of teacher-made tests “seems to indicate that training programs addressing item construction and tests as measurement of student learning are desirable” (p. 37).

What Experts Have Said Teachers Need to Know About Measurement and Student Assessment

Measurement specialists and educators have long voiced their views about what teachers need to know in the area of measurement. In 1964 Mayo conducted an extensive survey of teachers, principals and superintendents, college and university professors, and testing and research specialists. His purpose was to identify an ideal list of competencies for beginning teachers in the area of educational measurement.

Mayo’s (1967) survey results seemed to indicate that many respondents placed equal emphasis on teacher knowledge of standardized testing and classroom or teacher-made tests. Gullickson’s two
studies (1982, 1985) and the findings of Fleming and Chambers (1983) seemed to indicate that teachers rely most heavily on teacher-made tests for student evaluation and classroom instructional feedback. It would be very easy to conclude that measurement instruction for teachers should be concentrated on such areas as test construction, grading, item analysis, and establishment of reliability and validity. Fleming (1979) spoke to the issue of real-world classroom measurement: the routine use of teacher-made tests versus standardized tests to measure students’ learning. Although she agreed that standardized tests are not always indicative of material taught in the classroom, and that teacher-made tests may be preferable, she voiced clear concerns about the quality of teacher-made tests. She contended that the children in the classroom receive much more information about their learning from the teacher-made tests they routinely take than from standardized test results that usually do not affect student grades, and the results of which may never even be reported directly to the students. “Certainly the failure message is communicated much more frequently from the classroom test than the standardized test” (p. 5). Because of the possibility that failure messages are communicated to students due to faulty measurement instruments, Fleming proposed the following as classroom measurement needs in the 1980s, requiring the support of school districts:

1. There should be renewed efforts to improve preservice and inservice training in evaluation of instruction. Evaluation should be emphasized as a critical step within the teaching cycle.
2. There is a need for more effective and comprehensive training materials in educational evaluation.
3. There is a need to improve the operation of their district-wide measurement systems as a support to improvement of classroom measurement processes.

Additionally, Fleming identified the following needs in the area of instrumentation:

1. There is a need for improved teacher-made classroom tests at every level.
2. There is a need for assessment procedures which may be utilized within the emerging “new” models for teaching.
3. There is a need for improved procedures for measurement of writing.
4. There is a need for development of language assessment instruments for the support of bilingual programs in the schools.
5. There is a need to develop naturalistic methodology which has application to classroom assessment problems and which has utility for classroom teachers.
6. There is a need for developing options in criterion referenced measurement for the classroom teacher. (pp. 1-20)

The case presented by Fleming and Chambers and by Gullickson for concentration on teacher-made measurement is indeed strong. However, teachers, particularly elementary teachers, report using results of standardized tests (Gullickson, 1985). Rudman et al. (1980) provided some additional insights as to the reasons for emphasizing teacher knowledge about standardized tests. They indicated that teachers make critical decisions regarding student placement and programming early in the school year, and require information within the first 3 or 4 weeks of school in order to make such decisions. Many of these decisions are affected by results of standardized test scores available in the students’ files, as well as by teacher observations and intuition. Additionally, teachers may be responsible for the interpretation of standardized test scores to parents at parent-teacher conferences. Rudman et al. (1980) concluded that teachers need a variety of information sources in order to make appropriate decisions about grouping, placement, and instruction. Assessment and instruction should be incorporated in the classroom, and classroom teachers need the knowledge and skills to make this possible.

Other authors have attempted to identify measurement competencies needed by classroom teachers in broader terms. Robert Ebel (1962) developed the following principles of measurement for educational achievement:

1. The measurement of educational achievement is essential to effective education.
2. An educational test is not more or less than a device for facilitating, extending, and refining a teacher’s observations of student achievement.
3. Every important outcome of education can be measured.
4. The most important educational achievement is command of useful knowledge.
5. Written tests are well suited to measure the student’s command of useful knowledge.
6. The classroom teacher should prepare most of the tests used to measure educational achievement in the classroom.
7. To measure achievement effectively the classroom teacher must be (a) a master of the knowledge or skill to be tested and (b) a master of the practical arts of testing.
8. The quality of a classroom test depends on (a) the relevance of the tasks included in it, (b) the representativeness of its sampling of all aspects of instruction, and (c) the reliability of the scores it yields.
9. The more variable the scores from a test designed to have a certain maximum possible score, the higher the expected reliability of those scores.

10. The reliability of a test can be increased by increasing the number of questions (or independent points to be scored) and by sharpening the power of individual questions to discriminate between students of high and low achievement. (pp. 21-26)

Ebel's principles reflect an underlying agreement among the experts that measurement must be incorporated routinely into the instructional process. Farr and Griffin (1973) indicated that teachers need to be shown the close relationship between measurement and instructional decision making. They asserted it is perhaps too often the case that measurement is dealt with in the preservice education of teachers as an entity unto itself, with the result that "the basic principle underlying the discussion of what teachers need to know about measurement is that measurement should serve a purpose" (p. 19) is neglected. They developed the following "Outline of Measurement Concepts and Skills Needed by Classroom Teachers":

**Listing Instructional Decisions**
A. For which decisions can information be collected?
B. Which decisions require continuous information feedback and which require only periodic feedback?
C. Are the decisions consistent (valid) with a stated definition of the skills and behaviors to be taught?

**Developing Decision Alternatives and Determining Information Needs**
A. What are the measurable differences between alternatives?
B. What criterion[sic] are used to determine the feasibility of particular alternatives?

**Collecting Information**
A. How can information be collected validly and reliably?
B. What procedures are there for collecting information congruently with instruction?
C. What are the strengths and weaknesses of various data collecting procedures?
D. How can collected information be related to decision making?
E. How can teacher observations be made more valid and reliable?
F. How should teacher assessments be constructed? (p. 27)

Farr and Griffin believed this outline could serve as a guide in the development of teacher competencies in measurement that directly relate to the classroom behaviors of teachers.
3. DEVELOPMENT OF STANDARDS

The Status of Training Prospective Teachers in Student Assessment

Schafer and Lissitz (1987) conducted a survey of AACTE member institutions in an attempt to determine their requirements for education students in the area of measurement. Responses were received from 438 of 707 institutions. The authors reported that "with the exception of school counseling and special education programs, 49% or more of the programs surveyed do not require for certification a formal course in measurement" (p. 61). Many of the institutions suggested measurement is covered in other courses that are required in their programs, but the authors questioned the value of measurement being taught incidentally and/or by professors who lack specific expertise in measurement.

Roeder (1972) conducted a survey of 940 elementary school teacher training institutions nationwide. Based on 860 usable responses, the author made the following observations:

While only 270 institutions reported requiring prospective elementary classroom teachers to complete a course devoted exclusively to tests and measures, 470 institutions required a course in play activities and games. . . 633 institutions reported requiring courses in music methods for classroom teachers, and 637 institutions required one or more courses in the art methods for classroom teachers. (p. 240)

Gullickson (1985) noted that colleges often provide some instruction in measurement and evaluation, but the time devoted to such instruction is limited. He observed, "Each professor is likely to choose topics he or she perceives as most important to teachers. As such, the professor’s choices will depend upon his or her knowledge of measurement" (p. 96).

In reviewing the literature on teacher knowledge of measurement, Farr and Griffin (1973) reached the following conclusions:

1. There should be concern over the adequacy of teacher preparation in administering, scoring, and interpreting standardized tests for that part of the vital role that teachers seem to play in testing. Also, though teachers have only minimal coursework in measurement, what should be the content of a tests and measurements course is a vital question that pre-service and in-service educators must face.

2. Teachers do not know much about measurement concepts particularly in relation to normative data and standardized tests. What they should know in terms of measurement concepts is another critical question.
3. Most studies of teachers' measurement knowledge relate only to standardized tests and not classroom testing for planning instruction. Yet, teachers are using what knowledge they have of standardized tests to make critical decisions in regard to students' academic programs.

4. Teachers occupy a central role in the testing and evaluation process of their pupils. They are deeply involved in testing, standardized or otherwise.

5. Standardized achievement test scores about pupils are relied on heavily by teachers and could have important effects on teachers' attitudes and behaviors toward students, and might influence evaluations of classroom performance (e.g., Rosenthal Study [1968]). Teachers seem to have great faith in tests. (p. 23)

Rudman et al. (1980) published an extensive review of the literature on teacher preparation in assessment. In it they reported:

While there appears to be general agreement that teachers are not overly confident of their ability to interpret standardized test scores, the degree of confidence reported varies from researcher to researcher. Olejnik, (1979) in a study conducted among non-test specialists (counselors, teachers and building principals), found that over 90% of elementary and middle school educators indicated that they were at least "somewhat" confident of their ability to interpret test scores. The least confident were high school educationists. But when a mini-test similar to one given in college-level measurement courses was administered to the respondents, this self-reported "confidence" was not borne out. Most educationists correctly answered an item dealing with a percentile score (73%), yet a similar proportion missed an item that related norms to standards (77% incorrectly assumed that they were the same). They showed little understanding of the significance of stanine differences (only 35% recognized that a two stanine difference is significant), and very few could properly interpret a grade equivalent score (12%). On the basis of his study, Olejnik concluded that in spite of self-reported confidence it appeared that non-measurement specialists needed additional assistance in the interpretation of standard scores.

Stetz has conducted a series of studies aimed at determining the extent to which teachers and other educationists understand and accept standardized test results. His first study was a market survey of Stanford Achievement Test users (Stetz, 1977). Among a number of questions asked was one dealing with the types of scores they found most useful for assessment purposes. Both teachers and administrators reported that they preferred grade equivalents and percentile ranks for meeting their assessment needs; 59% of the teachers surveyed chose these two scores for individual student evaluation, 56% chose these two scores for class evaluation purposes, 65%
preferred these two scores for reporting test results to parents. One would like to assume from this that those who showed such a strong preference for these two standard scores understood what they signified, but Olejnik's study does give one some pause (Olejnik, 1979). (pp. 14-15)

Gullickson (1986) surveyed classroom teachers and professors responsible for teacher training to determine the measurement concepts viewed as important by the two groups. Gullickson reported strong disagreement between teachers and professors regarding statistics, nontest evaluation activities, and formative and summative evaluation:

Regarding statistics, two factors appear to be probable reasons for the teacher/professor disagreement. First, others who have assessed teachers' competency in measurement (see Rudman et al., 1980) have indicated that teachers do not have a good grasp of statistical concepts. This suggests that preservice measurement instruction, despite its relatively substantial emphasis on statistics, does not result in a level of understanding that would enable teachers to comfortably apply statistics to their evaluation needs. Such discomfort with statistics may well lead to devaluing of it.

Second, teachers may perceive such analyses as requiring more work than is justified by the benefits, particularly since statistical analyses can be avoided without obvious effect. In this regard, it is noteworthy that although there is substantial agreement among measurement experts as to the importance of statistical analyses, there is a paucity of empirical evidence to establish the positive instructional effects of such analyses.

Regarding nontest evaluation techniques, not only do professors give the topic substantially less emphasis than teachers recommend, but other research (Gullickson, 1985; Salmon-Cox, 1982; Stiggins & Bridgeford, 1982) indicates that teachers make substantial use of nontest evaluation techniques. Given their substantial use, greater emphasis on nontest evaluation techniques in preservice training programs should be expected. Here again there may be several reasons for the difference in professor and teacher opinions: (a) professors may not be aware of the extent to which teachers employ such techniques (research by Beck & Stetz, cited in Rudman et al., 1980, suggests that measurement experts do not have a clear understanding of teacher evaluation practices); (b) professors may perceive such techniques to be properly the domain of instructional methods courses and not the domain of measurement courses; and (c) professors may perceive the use of such techniques as less reliable and less valid than other evaluation techniques—thus deserving less emphasis.
Teacher and professor differences regarding formative and summative evaluation appear to stem from two possible roots. First, teacher priority items suggest that teachers recommend emphasis both on the general topics and on their specific applications. In contrast, professors give priority solely to the general issues with the expectation that specific applications will be provided in other methods courses. Certainly, given the diverse group of students who typically take an educational measurement course, presentation of examples appropriate to the needs of all students would be a time consuming and difficult task.

Second, five of the seven teacher-priority items for formative and summative evaluation relate directly to the identification and study of exceptional children (e.g., data to guide remediation, identifying gifted and slow learners, and identifying underachievers). None, however, was included among the professor priorities. This suggests that teachers alone place a high priority on the evaluation of special students. (pp. 350-353)

Perhaps Fleming (1979) addressed the teacher/professor conflicts most directly:

It appears that preservice teacher training with its emphasis on technical considerations and measurement processes as isolated events contribute to the ongoing dilemma for teachers. Is it too much to expect that training programs should foster a view of the instructional process as a continuum such as has been delineated by Tyler, for example, which in such a conceptualization consists of objectives, learning experiences and evaluation? (p. 2)

STANDARDS FOR TEACHER COMPETENCE IN EDUCATIONAL ASSESSMENT OF STUDENTS

By establishing standards for teacher competence in student assessment, the four involved associations subscribe to the view that student assessment is an essential part of teaching and that good teaching cannot exist without good student assessment. Training to develop the competencies covered in the standards should be an integral part of preservice preparation. Further, such assessment training should be widely available to practicing teachers through staff development programs at the district and building levels.

---

The committee that developed the standards represented four professional associations. James R. Sanders (Western Michigan University) chaired the committee and represented NCME along with John R. Hills (Florida State University) and Anthony J. Nitko (University of Pittsburgh). Jack C. Merwin (University of Minnesota) represented the American Association of Colleges for Teacher Education. Carolyn Trice represented the American Federation of Teachers. Marcella Dianda and Jeffrey Schneider represented the National Education Association. This section of the chapter represents the work of this committee and is a reproduction of the resulting document.
The standards are intended for use as:

- a guide for teacher educators as they design and approve programs for teacher preparation
- a self-assessment guide for teachers in identifying their needs for professional development in student assessment
- a guide for workshop instructors as they design professional development experiences for in-service teachers
- an impetus for educational measurement specialists and teacher trainers to conceptualize student assessment and teacher training in student assessment more broadly than has been the case in the past

The Approach Used to Develop the Standards

The memberships of the four associations are professional educators involved in teaching, teacher education, and student assessment. Members of these associations are concerned about the inadequate preparation of teachers for assessing the educational progress of their students, and thus sought to address this concern effectively. The committee named by the associations first met in September 1987 and affirmed its commitment to defining standards for teacher preparation in student assessment. The committee then undertook a review of the research literature to identify needs in student assessment, current levels of teacher training in student assessment, areas of teacher activities requiring competence in using assessments, and current levels of teacher competence in student assessment.

The members of the committee used their collective experience and expertise to formulate and then revise statements of important assessment competencies. Several drafts of these competencies were revised by the committee before the standards were released for public review. Comments by reviewers from each of the associations were then used to prepare this final statement.

Overview of the Standards

There were seven standards developed to cover assessment competencies needed by classroom teachers. In recognizing the critical need to revitalize classroom assessment, some standards focus on classroom-based competencies. Because of teachers’ growing roles in education and policy decisions beyond the classroom, other standards address assessment competencies underlying teacher participation in decisions related to assessment at the school, district, state, and national levels.
The scope of a teacher’s professional role and responsibilities for student assessment may be described in terms of the following activities. These activities imply that teachers need competence in student assessment and sufficient time and resources to complete them in a professional manner:

- **Activities occurring prior to instruction**: (a) understanding students’ cultural backgrounds, interests, skills, and abilities as they apply across a range of learning domains and/or subject areas; (b) understanding students’ motivations and their interests in specific class content; (c) clarifying and articulating the performance outcomes expected of pupils; and (d) planning instruction for individuals or groups of students.

- **Activities occurring during instruction**: (a) monitoring pupil progress toward instructional goals; (b) identifying gains and difficulties pupils are experiencing in learning and performing; (c) adjusting instruction; (d) giving contingent, specific, and credible praise and feedback; (e) motivating students to learn; and (f) judging the extent of pupil attainment of instructional outcomes.

- **Activities occurring after the appropriate instructional segment (e.g., lesson, class, semester, grade)**: (a) describing the extent to which each pupil has attained both short- and long-term instructional goals; (b) communicating strengths and weaknesses based on assessment results to students and parents or guardians; (c) recording and reporting assessment results for school-level analysis, evaluation, and decision making; (d) analyzing assessment information gathered before and during instruction to understand each student’s progress to date and to inform future instructional planning; (e) evaluating the effectiveness of instruction; and (f) evaluating the effectiveness of the curriculum and materials in use.

- **Activities associated with a teacher’s involvement in school building and school district decision-making**: (a) serving on a school or district committee examining the school’s and district’s strengths and weaknesses in the development of its students; (b) working on the development or selection of assessment methods for school building or school district use; (c) evaluating school district curriculum; and (d) other related activities.
Activities associated with a teacher's involvement in a wider community of educators: (a) serving on a state committee asked to develop learning goals and associated assessment methods; (b) participating in reviews of the appropriateness of district, state, or national student goals and associated assessment methods; and (c) interpreting the results of state and national student assessment programs.

Each standard that follows is an expectation for assessment knowledge or skill that a teacher should possess in order to perform well in the five areas just described. As a set, the standards call on teachers to demonstrate skill in selecting, developing, applying, using, communicating, and evaluating student assessment information and student assessment practices. A brief rationale and illustrative behaviors follow each standard.

The standards represent a conceptual framework or scaffolding from which specific skills can be derived. Work to make these standards operational will be needed even after they have been published. It is also expected that experience in the application of these standards should lead to their improvement and further development.

The Standards

1. Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.

Skills in choosing appropriate, useful, administratively convenient, technically adequate, and fair assessment methods are prerequisite to good use of information to support instructional decisions. Teachers need to be well acquainted with the kinds of information provided by a broad range of assessment alternatives and their strengths and weaknesses. In particular, they should be familiar with criteria for evaluating and selecting assessment methods in light of instructional plans.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be able to use the concepts of assessment error and validity when developing or selecting their approaches to classroom assessment of students. They will understand how valid assessment data can support instructional activities such as providing appropriate feedback to students, diagnosing group and individual learning needs, planning for individualized educational programs, motivating students, and evaluating instructional procedures. They will understand how invalid information can affect
instructional decisions about students. They will also be able to use and evaluate assessment options available to them, considering among other things, the cultural, social, economic, and language backgrounds of students. They will be aware that different assessment approaches can be incompatible with certain instructional goals and may impact quite differently on their teaching.

Teachers will know, for each assessment approach they use, its appropriateness for making decisions about their pupils. Moreover, teachers will know where to find information about and/or reviews of various assessment methods. Assessment options are diverse and include text- and curriculum-embedded questions and tests, standardized criterion-referenced and norm-referenced tests, oral questioning, spontaneous and structured performance assessments, portfolios, exhibitions, demonstrations, rating scales, writing samples, paper-and-pencil tests, seatwork and homework, peer- and self-assessments, student records, observations, questionnaires, interviews, projects, products, and others’ opinions.

2. Teachers should be skilled in developing assessment methods appropriate for instructional decisions.

While teachers often use published or other external assessment tools, the bulk of the assessment information they use for decision making comes from approaches they create and implement. Indeed, the assessment demands of the classroom go well beyond readily available instruments.

Teachers who meet this standard will have the conceptual and application skills that follow. Teachers will be skilled in planning the collection of information that facilitates the decisions they will make. They will know and follow appropriate principles for developing and using assessment methods in their teaching, avoiding common pitfalls in student assessment. Such techniques may include several of the options listed at the end of the first standard. The teacher will select the techniques which are appropriate to the intent of the teacher's instruction.

Teachers meeting this standard will also be skilled in using student data to analyze the quality of each assessment technique they use. Since most teachers do not have access to assessment specialists, they must be prepared to do these analyses themselves.

3. Teachers should be skilled in administering, scoring, and interpreting the results of both externally-produced and teacher-produced assessment methods.

It is not enough that teachers are able to select and develop good assessment methods; they must also be able to apply them properly.
Teachers should be skilled in administering, scoring, and interpreting results from diverse assessment methods.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be skilled in interpreting informal and formal teacher-produced assessment results, including pupils' performances in class and on homework assignments. Teachers will be able to use guides for scoring essay questions and projects, stencils for scoring response-choice questions, and scales for rating performance assessments. They will be able to use these in ways that produce consistent results.

Teachers will be able to administer standardized achievement tests and be able to interpret the commonly reported scores: percentile ranks, percentile band scores, standard scores, and grade equivalents. They will have a conceptual understanding of the summary indexes commonly reported with assessment results: measures of central tendency, dispersion, relationships, reliability, and errors of measurement.

Teachers will be able to apply these concepts of score and summary indices in ways that enhance their use of the assessments that they develop. They will be able to analyze assessment results to identify pupils' strengths and errors. If they get inconsistent results, they will seek other explanations for the discrepancy or other data to attempt to resolve the uncertainty before arriving at a decision. They will be able to use assessment methods in ways that encourage students' educational development and that do not inappropriately increase students' anxiety levels.

4. Teachers should be skilled in using assessment results in making decisions about individual students, planning teaching, developing curriculum, and school improvement.

Assessment results are used to make educational decisions at several levels: in the classroom about students, in the community about a school and a school district, and in society, generally, about the purposes and outcomes of the educational enterprise. Teachers play a vital role when participating in decision making at each of these levels and must be able to use assessment results effectively.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be able to use accumulated assessment information to organize a sound instructional plan for facilitating students' educational development. When using assessment results to plan and/or evaluate instruction and curriculum, teachers will interpret the results correctly and avoid common misinterpretations, such as basing decisions on scores that lack curriculum
validity. They will be informed about the results of local, regional, state, and national assessments and about their appropriate use for pupil, classroom, school, district, state, and national educational improvement.

5. Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.

Grading students is an important part of professional practice for teachers. Grading is defined as indicating both a student’s level of performance and a teacher’s valuing of that performance. The principles for using assessments to obtain valid grades are known and teachers should employ them.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be able to devise, implement, and explain a procedure for developing grades composed of marks from various assignments, projects, in-class activities, quizzes, tests, and/or other assessments that they may use. Teachers will understand and be able to articulate why the grades are rational, justified, and fair, acknowledging that such grades reflect their preferences and judgments. Teachers will be able to recognize and to avoid faulty grading procedures such as using grades as punishment. They will be able to evaluate and to modify their grading procedures in order to improve the validity of the interpretations made from them about students’ attainments.

6. Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.

Teachers must routinely report assessment results to students and to parents or guardians. In addition, they are frequently asked to report or to discuss assessment results with other educators and with diverse lay audiences. If the results are not communicated effectively, they may be misused or not used. To communicate effectively with others on matters of student assessment, teachers must be able to use assessment terminology appropriately and must be able to articulate the meaning, limitations, and implications of assessment results. Furthermore, teachers will sometimes be in a position that will require them to defend their own assessment procedures and their interpretations of them. At other times, teachers may need to help the public to interpret assessment results appropriately.

Teachers who meet this standard will have the conceptual and application skills that follow. Teachers will understand and be able to give appropriate explanations of how the interpretation of student assessments must be moderated by the student’s socioeconomic,
cultural, language, and other background factors. Teachers will be able to explain that assessment results do not imply that such background factors limit a student’s ultimate educational development. They will be able to communicate to students and to their parents or guardians how they may assess the student’s educational progress. Teachers will understand and be able to explain the importance of taking measurement errors into account when using assessments to make decisions about individual students. Teachers will be able to explain the limitations of different informal and formal assessment methods. They will be able to explain printed reports of the results of pupil assessments at the classroom, school district, state, and national levels.

7. Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

Fairness, the rights of all concerned, and professional ethical behavior must undergird all student assessment activities, from the initial planning for and gathering of information to the interpretation, use, and communication of the results. Teachers must be well versed in their own ethical and legal responsibilities in assessment. In addition, they should also attempt to have the inappropriate assessment practices of others discontinued whenever they are encountered. Teachers should also participate with the wider educational community in defining the limits of appropriate professional behavior in assessment.

Teachers who meet this standard will have the conceptual and application skills that follow. They will know those laws and case decisions that affect their classroom, school district, and state assessment practices. Teachers will be aware that various assessment procedures can be misused or overused, resulting in harmful consequences such as embarrassing students, violating a student’s right to confidentiality, and inappropriately using students’ standardized achievement test scores to measure teaching effectiveness.

WHERE DO WE GO FROM HERE?

In 1986, after studying the feasibility of the NCME taking on a licensing or certifying (i.e., credentialing) role for measurement experts, it was noted that the nature of measurement expertise in education was too illusory ever to be able to define, or standardize, requirements across the education profession. Instead, collaborative studies with professional education associations were planned to
identify the assessment competencies needed to perform in different professional roles, and to prepare joint statements about the preservice and in-service preparation in student assessment of educators filling these different roles.

The classroom teacher role was the first to be studied. The resulting standards are intended to be a statement that will affect teacher certification requirements and the accreditation of teacher preparation programs. There is an expectation that administrator, counselor, testing director, special education director, curriculum director, and other roles will require similar attention in the future.

Now that the teacher standards have been developed, there are a number of follow-up activities that deserve the attention of the four collaborating associations. These include:

• collaborating on a table of specifications for each standard, and then developing assessment procedures and instruments for assessing the extent to which an individual can meet the standards.
• collaborating on instructional modules and workshops for teachers based on the standards.
• collaborating on developing a curriculum strand to prepare preservice teachers for student assessment. This curriculum strand might contain grounded scenarios of classroom teaching in which teachers are meeting and not meeting the standards, with analyses and instruction to accompany each scenario.
• collaborating on the dissemination and use of the standards through the four associations, state departments of education, and such projects as the National Board of Teaching.

Another thrust for the future would be for the NCME to work with the American Association of School Administrators (AASA), the National Association of Secondary School Principals (NASSP), and the National Association of Elementary School Principals (NAESP) to prepare similar standards for school administrators. This pattern of collaborative development could then continue for educator groups that include testing directors, counselors, special education specialists, curriculum specialists, and other professional groups that might be added. By the time standards and spinoff products are developed and are being used for each of these groups, it would then be time to review and update each set of standards in a collaborative and systematic manner. A review by the cooperating associations every 5 years would be in order.
There is still a great deal of work to be done to improve the quality of student assessments in education. The first step taken by the four associations to develop these standards for teacher competence in student assessment is a major step in the right direction.

REFERENCES


