THE NEED FOR STUDENT EVALUATION STANDARDS

Prepared by

The Joint Committee on Standards for Educational Evaluation

Arlen R. Gullickson
Chair

May 2000

This paper is taken in large part from the Introduction to the first draft of *The Student Evaluation Standards*. Permission to copy, distribute, and publish this paper is granted with the proviso that credit for the article is given to the Joint Committee on Standards for Educational Evaluation. The following persons participated in preparing this paper: Jennifer Fager, Dale Farland, Arlen Gullickson, and James Sanders.
The Joint Committee on Standards for Educational Evaluation (JC) is currently developing student evaluation standards intended to serve primarily in classroom settings across the broad array of elementary, secondary, and postsecondary settings. The need for these standards is found both in the JC’s mission and in concrete findings from reports and research studies on this topic. This paper briefly addresses these identified needs as a means to increase understanding of why the JC has chosen to undertake this substantial effort.

The JC began its work in 1975 and has developed two sets of standards that currently serve the educational field. These standards, *The Program Evaluation Standards* (1994) and *The Personnel Evaluation Standards* (1988), address two of the three primary ingredients in classroom instruction: the curriculum and the instruction provided by the teacher and others. The remaining and most important ingredient, the students themselves, is addressed by the new standards. The purpose of the standards is to provide direction to teachers and others to ensure that their evaluations of students provide optimal guidance to the students, serving their educational development needs as well as serving the information needs of others.
The expressed purposes of the student evaluation standards, like those of the previously developed standards flow directly from the mission of the Joint Committee. “The mission of the reconstituted Joint Committee on Standards for Educational Evaluation is to promote concern for evaluations of high quality based on sound evaluation practices and procedures, and to meet existing and emerging needs in the field of evaluation” (Operating Procedures for the Development and Approval of Standards of the Joint Committee on Standards for Educational Evaluation, 1988). Support for this mission is provided by the stated rationale for the Joint Committee in the same document:

1. Lifelong education is important to individuals, institutions within society, and society itself.
2. Evaluation is an integral, inevitable feature of all aspects of education.
3. Evaluations that are properly commissioned, conducted, disseminated, and used help improve education in general, and specific programs, products, and materials in particular.
4. Evaluation standards can play a vital role in upgrading and ensuring the quality of educational evaluations.

Additional compelling reasons for developing the student evaluation standards are
provided through study of student evaluations as conducted by teachers across a wide variety of settings. The remainder of this paper looks directly at these reasons as a needs basis for the student evaluation standards. First, and most directly, we know that student evaluations exist and persist in our schools. Secondly, we know that there are substantial barriers to sound student evaluations.

The significant role of student evaluations. Sound student evaluation is an essential ingredient to strong educational programs. Evaluation is probably the most common and pervasive aspect of student instruction. It is the primary tool for guiding student development, crossing all academic disciplines. Certainly, evaluations of students occur in all classrooms and regularly confront students and educators in a wide variety of decision situations that affect their educational development. These decisions include matriculation, admissions, grading, tracking, instructional decisions for individual students, discipline, and merit awards.

Sound student evaluations also serve many other information needs such as employment decisions, evaluations of teachers, progress, school programs, school policies, and public accountability. For example, carefully prepared portfolios of student evaluations information can significantly serve students in making the case that they have employable skills and can assist potential employers in determining whether to hire the student. Also, ensuring the academic growth and well-being of students are
The Need for Student Evaluation Standards

teacher duties. The evaluations of students (e.g., aggregated findings across students and disaggregated by student characteristics) can provide substantial information regarding the performance of teachers.

When student evaluations are not sound, educational programs suffer. Poor student evaluations victimize and harm students. When questionable evaluation practices are employed in high-stakes situations (i.e., those conducted to serve important decisions and those substantially impacting students), the negative consequences are likely to draw attention. In addressing the effects of poorly conducted large-scale evaluation activities, several authors have noted their concerns about the increase in the number of high-stakes decisions based on limited student information (Brandt, 1989; Campbell & Greenberg, 1993; National Forum on Assessment, 1995; Pollard, 1993; Shepard, 1989).

While highly visible high-stakes evaluations capture attention, far more evaluations touch individual students on a daily basis. These more obscure evaluations, when poorly conducted, can be equally devastating for individual children. For example, students, their parents, and others use information from evaluations to set goals and expectations. If the evaluation information is incorrect or unsound or if evaluative conclusions are unjustified, individuals’ decisions are likely to be academically, economically, and socially harmful to the student's growth.
The Need for Student Evaluation Standards

Educators need to be certain they are conducting appropriate evaluations for each student and that the results given to the student and others are accurate. In addition, if the evaluations used by classroom teachers are conducted by outsiders, the teachers need to be good consumers and judges of these evaluations.

**Barriers to sound evaluation.** Unfortunately, there is ample evidence that sound student evaluation is not the norm in our schools. There is further evidence that the quality of student evaluations is not likely to change without major, systematic efforts. Some of the most critical impediments to sound student evaluation and to meaningful change in evaluation practices include those listed below:

1. **Inadequate Teacher Preparation**  Studies conducted since the early 1960s have consistently shown that teachers are not well prepared to conduct student evaluations (Gullickson, 1993; Mayo, 1967; Sanders & Vogel, 1993; Shafer & Lissitz, 1987; Stiggins, 1991; Stiggins & Conklin, 1992; Wolmut, 1988). Nearly half of teacher education institutions provide no preparatory coursework in student evaluation or classroom assessment practices. In those that do, most instruction and supporting materials focus on testing, with little attention to how to organize, conduct, and use these assessments in the instructional settings they are intended to serve. As a result, teachers do not have adequate preparation to properly evaluate students, nor do they have strong support
materials to guide their efforts.

2. **Inadequate Technical Support** To further complicate this issue, there is minimal technical assistance available to educators because most administrators and support personnel also lack training in student evaluation (Gullickson, 1993).

3. **Changing Demands** Student evaluation practices need to respond to ever-changing policies and situations in education (Crooks, Kane, & Cohen, 1996; Shepard, 1989). Educators constantly face the need to select and modify evaluation systems to meet the needs of all audiences for student evaluation, regardless of abilities or development level (Brandt, 1989; Gallagher, 1995; Teeter, 1995; Wagner, 1995). This adaptability is especially challenging when teachers are not well prepared in evaluation techniques.

4. **Professional Disagreement** Professionals disagree on many evaluation practices (e.g., grading of students), leaving teachers on their own to determine appropriate practices. In the case of grades and grading, the resulting difficulties are almost legend. For example, Kelley (1989), Hills (1991), and Wiggins (1994) identified problems with grade validity and reliability due to grade inflation, inconsistent and often biased practices by different evaluators, and the use of
criteria that are not valid for grading purposes. As a result, grading practices vary so significantly across teachers and institutions that grades have limited value as an indicator of student learning.

5. **Lack of Shared Language** Too often the message of evaluation and its importance to student learning and growth are not communicated clearly. In part, this may be due to language barriers. Evaluators and measurement specialists use terminology in ways that are different from the language typically used by teachers, administrators, students, and parents. The resulting miscommunications can lead to inappropriate practice even by those seeking to use evaluations appropriately. The language dilemma confounds communication both ways, as evaluation and measurement specialists too often don’t have a very clear understanding of the day-to-day realities of the K-12 classroom.

6. **Inadequate School Policy** Schools too often fail to develop and provide sound policies to guide evaluations of students. Anecdotal reports from teachers suggest that policies are sometimes at odds with fair and accurate evaluations of students, placing teachers and their students in a Catch-22 situation. Grading again serves as one example: There are still school districts that require grading on a fixed curve despite strong arguments and evidence that it is inappropriate.
7. **Transitory Interventions** The adoption of change literature (Rogers, 1995) suggests that change doesn’t happen quickly, pervasively, or thoroughly; nor can it be assumed that changes will endure. Educators over the past several decades have gotten used to “reform movements” and new programs. Unless they are convinced that there are real benefits to them and/or their students, they put up with the change only as long as the program is being sponsored by the administration or other agency. A quote from an elementary teacher captures the too-frequent reaction, “After wasting a lot of time on the program, we’re all ready to go back to the way we were doing things. In a year, you will see little evidence of it. We just don’t have the time to keep doing all the extra things required.”

8. **Ineffective Instruction** Few undergraduate or graduate educational measurement courses would be voted as the “Best Liked Course” by college students. Whether deserved or not, learning more about educational measurement or evaluation is not one of the highest priorities expressed by teachers. The result is that teachers have generally tried to avoid the very instruction in which they would be most likely to acquire skills for engaging in sound student evaluation.

**Removing the barriers.** Findings such as those above have led several authors
(Conley, 1997; Gronlund, 1976; Joint Advisory Committee, 1993) to argue that there is a need for a common language in student evaluation. A better understanding of the best practices required of evaluators is necessary to avoid common errors that can be detrimental to students and consumers of student evaluations. Others (Anderson, 1985; Cochran, 1995; Webster, 1995) have directly argued that standards are needed to guide evaluative decisions about students, develop grading systems that reflect appropriate criteria, and improve preparation of those who conduct and use evaluation.

As is true in personnel evaluation and program evaluation (Joint Committee on Standards for Educational Evaluation, 1988, 1994), evaluation professionals believe that there are common themes that are knowable, teachable, and applicable across many student evaluation situations. For example, few would argue the importance of gathering reliable information to conduct student evaluations. The challenge is to develop guidelines that promote sound practice without prescribing specific techniques that may not serve well in all situations or across time.

The Joint Committee’s sponsoring organizations have worked to develop standards

that guide the use of assessments, particularly tests. These standards provide strong support for helping to assure that assessments, especially high-stakes assessments are well constructed and provide consistent, accurate, and meaningful information to serve in student decision situations. Textbooks provide substantial information about the construction of tests and other evaluation tools, again to help teachers and others properly develop and employ these tools. These previous efforts all focus on tools of evaluation. Though useful, they do not provide clear guidance for the myriad ways in which tools, procedures, human intuition and judgment, and the politics of the classroom must work together in the evaluation process to provide feasible, strong, defensible, and accurate evaluation support to the teachers and students.

The Joint Committee’s efforts to develop student evaluation standards marks the first attempt to create principles of sound student evaluation-standards—that not only focus on the tools employed but on the human elements and the contextual settings. For example, there are always strong trade-offs between construction of highly reliable instruments and feasibility. Often, the actual costs of construction and validation may preclude use of some types of instruments. Yet, teachers must make evaluative judgments to guide students regardless of the quality of available data.

Additionally, only rarely do evaluations of students in the classroom depend on only
one source of data. For example, student grades often reflect student performance across an array of homework, in-class participation, exams, reports, etc. Often, parents must be provided with information, and steps must be taken to ensure that the information is helpful and supportive of student development. School demands for teacher time regularly mitigate against well-developed assessments. In these situations, teachers must know what to do and how to proceed. All of these factors argue for standards that help to identify when the evaluation tasks have been conducted properly, but do constrain teachers and others to specific practices.

The Joint Committee’s student evaluation standards recognize that student evaluations are integral to student development and informing education stakeholders, but also recognize the many factors cited above that work against sound evaluations. We know that evaluations can carry substantial risks for students, especially when they are improperly or poorly conducted. As such, the standards are intended to help education professionals (teachers, principals, teacher educators, etc.) recognize and put into practice these important principles and practices, so that the profession can move ahead to breach the barriers and ensure that students receive evaluations helpful to their educational and personal development.

Student evaluation standards are available in “first-draft form.” They are currently undergoing review by national and international review panels. You can access, read,
review, and provide input to the standards development process by visiting the Joint Committee web site at (jc.wmich.edu). That site also provides an opportunity for you to assist in the coming field-test process for these standards. Please take this opportunity to become involved in this important task.
References


In T. R. Guskey (Ed.), *Communicating student learning: Association for Supervision and Curriculum Development yearbook*. Alexandria, VA: ASCD.
