

# Improving Schools

<http://imp.sagepub.com/>

---

## Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development

Craig A. Mertler

*Improving Schools* 2009 12: 101

DOI: 10.1177/1365480209105575

The online version of this article can be found at:

<http://imp.sagepub.com/content/12/2/101>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Improving Schools* can be found at:**

**Email Alerts:** <http://imp.sagepub.com/cgi/alerts>

**Subscriptions:** <http://imp.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://imp.sagepub.com/content/12/2/101.refs.html>

>> [Version of Record](#) - Jun 23, 2009

[What is This?](#)

# Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development

**Craig A. Mertler**

*University of West Georgia, USA*

---

## **Abstract**

Assessing student performance is one of the most critical aspects of the job of a classroom teacher; however, many teachers in the United States do not feel adequately prepared to assess their students' performance. These feelings of inadequacy are exemplified when placed against the context of the No Child Left Behind Act of 2001, especially with its increased focus on accountability and assessment. This study examined the effectiveness of a two-week classroom assessment workshop for inservice teachers. The workshop was based on the *Standards for Teacher Competence in the Educational Assessment of Students* and focused on discussion, practice, and practical application through performance assessment tasks. The study utilized a parallel mixed-methods design. Teachers were pre-tested and post-tested using the *Assessment Literacy Inventory*. Additionally, teachers kept daily reflective journals in order to document their experiences. The training was shown to be highly effective for the teachers, as evidenced through the dramatic increase in post-test scores over pre-test scores, and perhaps even more so through critical examination of their reflective journals.

**Keywords:** assessment, professional development, teacher competence, teacher perceptions, professional reflection

## **Introduction**

Assessing student performance is one of the most critical aspects of the job of a classroom teacher. It impacts nearly everything that teachers do. However, many teachers in the United States do not feel adequately prepared to assess their students' performance (Mertler, 1998, 1999; Stiggins, 1999). They often believe that the assessment training that they received as undergraduates did not prepare them to be comfortable with the decisions they are routinely charged to make. Studies have shown that teachers' *feelings* of discomfort and inadequate preparation are in fact mirrored by their limited 'assessment literacy' (Popham, 2003). These facts are disconcerting in and of themselves; however, when placed in juxtaposition with the No Child Left Behind Act of 2001, the relative importance of this apparent inadequacy is substantially heightened. This is largely due to the fact that the law has increased the national focus on assessment and accountability, at the classroom, district, and state levels.

When considering teachers' levels of assessment preparation, Plake (1993) found that over 70 percent of teachers responding to a national survey reported exposure to tests and measurement content (either through a course or inservice training), although for the majority it had been longer than six years. Inservice teachers who had previous coursework/training scored higher on a test of assessment literacy than those who had not had such coursework; although the difference in scores was less than one point, it was statistically significant.

Teachers often claim that their lack of preparation is largely due to inadequate preservice training in educational measurement (Plake, 1993). For example, in a statewide survey asking inservice teachers about their perceived level of preparedness to assess student learning resulting specifically from their teacher preparation programs, over 85 percent of the respondents reported that they were not well prepared (Mertler, 1999). When asked about their current level of preparedness, slightly more than half indicated that they were well prepared to assess student learning. Mertler (1999) concluded that this potentially implies that teachers tend to develop assessment skills on the job, as opposed to structured environments such as courses or workshops.

It should not come as a surprise that, in this age of No Child Left Behind (NCLB) and Adequate Yearly Progress (AYP), issues related to classroom assessment continue to be a growing concern for educators. Even as far back as 15 years ago, much federal funding in the United States was contingent on measured student performance, as reported at the district level. As the decade of the 1980s progressed, greater emphasis was placed on the quality of student assessment at the classroom level. A need developed for a set of principles to guide what teachers should know and be able to do with respect to classroom assessment. The *Standards for Teacher Competence in the Educational Assessment of Students* (AFT, NCME, & NEA, 1990) were a joint effort between the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association. They were originally developed in order to address the problem of inadequate assessment training for teachers (AFT, NCME, & NEA, 1990). *The Standards for Teacher Competence in the Educational Assessment of Students* specifies teacher skills in the following areas: Choosing and Developing Assessment Methods; Administering, Scoring, and Interpreting Assessment Results; Using Assessment Results for Decision Making and Grading; Communicating Assessment Results; and Recognizing Unethical Assessment Practices.

The term 'assessment literacy' was alluded to earlier in this section. 'Assessment literates', as described by Stiggins (1995), 'know the difference between sound and unsound assessment. They are not intimidated by the sometimes mysterious and always daunting technical world of assessment' (p. 240). He notes that assessment-literate educators (regardless of whether they are teachers, administrators, or superintendents) enter the realm of assessment knowing what they are assessing, why they are doing it, how best to assess the skill/knowledge of interest, how to generate good examples of student performance, what can potentially go wrong with the assessment, and how to prevent that from happening. They are also aware of the potential negative consequences of poor, inaccurate assessment (Stiggins, 1995).

The *Standards* acknowledge and specify the importance of teacher education and professional development in the area of classroom assessment (Brookhart, 2001). Numerous research studies have been conducted over the past decade or so that have

addressed one or more of the seven *Standards* (Brookhart, 2001). However, only a handful of studies have addressed the *Standards* in their entirety: one study focused on the application of the *Standards* to inservice teachers (Plake, 1993); two studies did the same for preservice teachers (Campbell et al., 2002; Mertler and Campbell, 2005); and a fourth measured assessment literacy for both inservice and preservice teachers (Mertler, 2003a).

In 1991, a national study was undertaken in order to measure teachers' assessment literacy (Plake, 1993). The *Standards* were used as a test blueprint for the development of the survey instrument (titled the *Teacher Assessment Literacy Questionnaire*) to be used in the study. A representative sample from around the United States was selected to participate; a total of 98 districts in 45 states surveyed, yielding a total usable sample of 555 respondents (Plake, 1993). It was concluded in the study that teachers were not adequately prepared to assess student learning, as evidenced by the average score of 23 (66%) of 35 items answered correctly.

A similar study, conducted by Campbell et al. (2002), attempted to apply the identical previously described assessment literacy instrument to undergraduate preservice teachers. The renamed *Assessment Literacy Inventory* was administered to 220 undergraduate students following completion of coursework in tests and measurement. The preservice teachers ( $M = 21$ ) averaged two fewer questions answered correctly than did the inservice teachers ( $M = 23$ ) in the Plake (1993) study.

Mertler (2003b) studied the assessment literacy of both preservice and inservice teachers, and then statistically compared the two groups. Using a slightly modified version of the *Teacher Assessment Literacy Questionnaire*, he obtained similar results to both the Plake et al. (1993) and Campbell et al. (2002) studies. The average score for inservice teachers was equal to 22 items answered correctly – quite similar to the average score of 23 obtained by Plake (1993). The average score for the preservice teachers was equal to 19 – also similar to the average score obtained by Campbell et al. (2002). The similarity of the results lent credibility to the notion that teachers – both preservice and inservice – do not seem to possess high, or perhaps even adequate, levels of assessment literacy.

It is interesting to note that both the Campbell et al. (2002) and the Mertler (2003b) studies were in essence replications of the Plake (1993) study, in that both used the same original instrument developed by Plake. Both Campbell et al. (2002) and Mertler (2003b) recommended a complete revision and/or redevelopment of the assessment literacy instrument. Campbell and Mertler undertook a study which focused on the development of a new, contextualized version of the *Assessment Literacy Inventory*, or *ALI* (Mertler and Campbell, 2004, 2005). These studies examined the assessment literacy of preservice teachers who had just completed a course in applied classroom assessment. Furthermore, explicit efforts were made to link course content, assignments, and experiences characteristic of educational decisions and practices outlined in the *Standards*. However, similar to previous studies of preservice teachers, those in these studies answered approximately 68 percent ( $M = 28.83$  out of 35) of the items correctly. When examining preservice teachers' overall performance on the *ALI*, it should be noted that their mean score was far lower than might otherwise be expected given their recent completion of coursework in classroom assessment. The researchers hypothesized that the observed gap between the preservice

teachers' *ALI* performance and recent formal training may be related to preservice teachers' limited classroom experience. Perhaps because the *ALI* was designed to test application of assessment concepts, limited familiarity with the day-to-day realities of the classroom may have precluded preservice teachers from making necessary connections.

The majority of previous studies on the topic of teachers' assessment knowledge and practices have tended to focus on what teachers *do not* know and what they *cannot* do in the classroom with respect to assessment. In one of the earlier studies based on the *Standards*, O'Sullivan and Johnson (1993) examined the impact that a graduate-level course in measurement could have on teachers' assessment competencies, according to the *Standards*. They reported a statistically significant increase in the teachers' scores on a measure of assessment competencies. The researchers attributed this increase in assessment competence to the performance-based nature of the course.

The study at hand is an attempt to examine what teachers know and are able to do following an intensive two-week training in classroom assessment.

### **Purpose of the study**

The purpose of this study was to examine the effectiveness of a two-week workshop for inservice teachers on the topic of classroom assessment. The workshop focused on the improvement of the teachers' classroom assessment knowledge and skills through the use of discussion, practice, and performance assessment tasks, focusing on the practical application of those knowledge and skills. The specific research questions being investigated were:

- What impact does a two-week workshop on the topic of classroom assessment have on teachers' assessment literacy?
- What are teachers' impressions of the assessment training with respect to its applicability and relative impact?

### **Methods**

A two-week classroom assessment workshop (titled 'Assessment Literacy for Teachers') was taught to a small group ( $n = 7$ ) of inservice teachers. All of the participating teachers had classroom experience at the elementary level and represented five different school districts. It is important to note that none of these teachers had ever received formal education on the topic of classroom assessment in the form of a stand-alone course, either in their preservice or graduate programs. The workshop was intended to help inservice teachers develop a deeper understanding of the process of assessing and evaluating student academic performance as it specifically relates to the instructional process. Furthermore, the participants were to develop skills in the construction and use of assessment instruments, and the interpretation of assessment results, for purposes of formative and summative evaluations, placement, and diagnosis. Seven teachers participated in the workshop.

Major topics covered during the workshop included the following: norm- and criterion-referenced measurements, validity and reliability of assessments, the integration of teaching and assessment, construction and use of traditional assessments, construction

and use of authentic assessments, grading, and the interpretation of standardized test results. The participants were assessed through their accomplishments on several performance assessment tasks, designed to parallel the *Standards for Teacher Competence in the Educational Assessment of Students* (AFT, NCME, & NEA, 1990). These nine performance assessments were:

- TASK #1 – Identifying & reviewing published tests
- TASK #2 – Developing an objective test
- TASK #3 – Summarizing objective test results
- TASK #4 – Conducting an item analysis
- TASK #5 – Determining validity and reliability of objective tests
- TASK #6 – Interpreting & communicating standardized test results
- TASK #7 – Developing valid grading procedures
- TASK #8 – Developing a performance assessment & scoring rubric
- TASK #9 – Determining unethical assessment practices

This study utilized a parallel mixed-methods design (Mertens, 2005; Tashakkori and Teddlie, 1998), using what Creswell (2003) calls a ‘concurrent triangulation strategy’, where both quantitative and qualitative data are collected simultaneously and analyzed in a complementary manner. The two types of data are collected independently *and* at the same time. Furthermore, the final conclusions and inferences are based on the results of analyses of both types of data (Mertens, 2005).

Teachers were pre-tested on the first day and post-tested on the last day of the workshop using the *Assessment Literacy Inventory* (Mertler and Campbell, 2004, 2005), which was also developed to parallel the *Standards for Teacher Competence in the Educational Assessment of Students*. Pre-test and post-test mean scores were informally compared (since the small sample size prohibited the use of inferential statistical analyses). Additional quantitative comparisons were done on the mean scores for each subscale, corresponding to each of the seven *Standards*.

Teachers were also asked to keep daily reflective journals on their experiences, both prior to and as a result of the workshop. Teachers were required to make an initial entry in the journal regarding a ‘personal inventory’ of their assessment knowledge. Each day, they were required to reflect on their knowledge of that particular day’s topic, both prior to and following the day’s lesson. A final journal entry gave them the opportunity to reflect on the influence of the training as a whole. Daily reflective journal entries were analyzed by means of content analysis in an attempt to examine similarities and trends across the participants and assessment topics.

## Results

The two-week inservice training in classroom assessment was shown to have a substantial impact in terms of improving teachers’ knowledge and understanding of various concepts and processes related to the assessment of student academic performance. The teachers’ performance on the 35-item *Assessment Literacy Inventory* post-test ( $M = 28.29$ ) was nearly nine points higher than their performance on the pre-test ( $M = 19.57$ ). The greatest amounts of improvement were experienced on Standard 5 – *Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments* (with the mean increasing from 2.43 to 4.43 on a five-point scale) and Standard 2 – *Teachers should be skilled in developing assessment methods appropriate for instructional*

**Table 1:** Mean pre-test, mean post-test, change scores, and ranking by standard on the *Assessment Literacy Inventory*

Scale	Pre-test mean	Post-test mean	Change (+/-)	Rank
Standard 1 <sup>a</sup>	3.29	3.43	+1.14	7
Standard 2 <sup>a</sup>	2.00	3.86	+1.86	2
Standard 3 <sup>a</sup>	2.71	3.86	+1.15	4
Standard 4 <sup>a</sup>	3.00	4.00	+1.00	5
Standard 5 <sup>a</sup>	2.43	4.43	+2.00	1
Standard 6 <sup>a</sup>	2.57	4.29	+1.72	3
Standard 7 <sup>a</sup>	3.71	4.57	+0.86	6
Total <sup>b</sup>	19.57	28.29	+8.72	

Note:  $n = 7$ .

<sup>a</sup>Each subscale (corresponding to a *Standard*) was composed of a maximum possible value equal to 5 points.

<sup>b</sup>The total score had a maximum possible value equal to 35 points.

*decisions* (with an increase from 2.00 to 3.86). A summary of the subscale and total scores for the pre-test and post-test is provided in Table 1.

Not only did the overall mean improve from pre-test to post-test, but *each* of the seven teachers improved their individual scores from pre-test to post-test. Scores for four of the seven teachers improved by at least 10 points on the total scale of 35 possible points. The individual teachers' performances are summarized in Table 2.

Further evidence of the impact of the assessment training was provided in the teachers' reflective journals. The results of the content analysis of journal entries have been organized by day (since each day's entry pertained to a specific assessment topic).

### *DAY 1 – Inventory of assessment knowledge (& Task #1)*

First and foremost, the teachers admitted that their knowledge about assessment was limited. The teachers seemed to have some basic knowledge about what assessment is

**Table 2:** Individual teachers' pre-test, post-test, and change scores on the *Assessment Literacy Inventory*

Teacher	Pre-test score <sup>a</sup>	Post-test score <sup>a</sup>	Change (+/-)
Teacher 1	22	28	+6
Teacher 2	19	26	+7
Teacher 3	19	30	+11
Teacher 4	15	26	+11
Teacher 5	19	29	+10
Teacher 6	22	28	+6
Teacher 7	21	31	+10

<sup>a</sup>The total scores for both the pre-test and post-test had a maximum possible value equal to 35 points.

and what it is not. Most tended to associate assessment with formal pencil-and-paper or standardized tests. Two of the teachers stated the following:

Assessment to me has always been 'tests' – unit tests, spelling [tests], etc. [combined to equal] student grades.

When I think of classroom assessment, paper and pencil tests [are] what generally comes to mind.

All of the teacher participants stated that they knew that assessment is more than just teaching and giving tests (for grades), but several were unable to describe various other means for assessing students. Those who could, described assessment as being formal (e.g. standardized tests) or very informal (e.g. observation and questioning).

### *DAY 2 – Objective and Subjective Tests Items (& Task #2)*

The focus of this day was the development of objective (i.e. multiple choice, alternate choice, and matching) and subjective (i.e. short answer and essay) test items. The teachers approached this topic with a certain amount of confidence. They all reported that they had been writing tests for some time; however, they came away from the class realizing how limited their knowledge and experiences had actually been. One teacher explained,

I have been teaching for 7 years and have made tests (many) throughout each school year. Up to this point I have not always been aware of the importance of the various formats for testing and what role they play in a student's learning. As I reflected on my own teaching practices, I have realized that I do not use a variety of testing formats with my students and need to improve on that area of my teaching . . . I have learned so much about the various types of [items]. After today's discussion I now realize that my tests are not as good as they should be . . . I realized that I do mostly (if not all) objective test. Next fall, I look forward to chang[ing] the way I construct my tests and try[ing] to include more subjective tests in my assessments.

The teachers seemed to be very surprised about various guidelines for writing objective and subjective written test items, including the development of a table of specifications:

I never thought about what should 'really' go into making a multiple choice test – t/f or matching . . . A lot of things I never thought about . . . I think the table of specification[s] is a great idea . . . Bloom's has always played a large part in my lesson planning – now I see how it fits into the objective tests.

Comments from other teachers reiterated this notion, along with the fact that writing good test items is not an easy task:

I am attempting to write a totally different type of test for me. Most of my test are either match[ing] or essay . . . I found myself questioning what I am doing numerous times.

I learned a lot of new things about developing objective and subjective tests. Some things I've been doing wrong all along – I've already gained several good ideas . . . I was very conscience of the guidelines for writing objective tests tonight.

Writing the test questions were harder than I expected! I now realize how easy it is to write test questions – but to write test questions *correctly* [emphasis added] is much harder! I was second-guessing the wording in my test questions and my distractors [sic].

Writing the test items was difficult. Since beginning this class, I was either reminded or learned about what to consider when creating an objective test. Having those things so fresh in my mind made it difficult because I was trying to apply them. It was quite interesting (intriguing).

I have really learned what goes into making a 'good' test – it can't be done the night before; it takes work.

### *DAYS 3 & 4 – Descriptive and Item Analyses (& Tasks #3, #4)*

The teachers seemed to be in complete agreement with each other in regard to their anxious anticipation of statistically analyzing their test results. They were not sure about what to expect, since none of them had ever done this before.

Today we are going to work with item analysis. I freely admit tons of confusion. I guess, once I actually look at the data and we discuss it, I'll feel better.

I'm a little frightened of all the statistics formulas.

However, once the teacher participants attempted both descriptive analyses and item analyses on their own test items, they realized that it was something that they *could* do and that it did, in fact, provide them some insight into the quality of their test items.

The test analysis outcome was interesting. The information discussed in class was understandable now that I am able to apply it.

Developing the item analysis was interesting for me. It gave me a lot of information about my test. I went back and made some changes to it which I think would help. It would be interesting to re-administer and calculate to see if there is a difference. Very intriguing!

I feel like I'm actually learning what some of these numbers mean.

While it was informative to statistically determine how reliable the whole test was, I found it much more helpful to look at each item on the test.

I'm fairly proud of my ability to use [the statistical software program].

While all of the teachers expressed an improved understanding of the whys and hows of analyzing classroom test data, many still had questions or continued to find some aspect of the process confusing. One teacher did not seem to be daunted by this, as she anticipated applying what she had learned:

It still seems confusing to me as far as the interpretation. Hopefully, I'll be able to calculate these things for at least a few real tests next school year. I'm very curious!

Another teacher stated:

Things are starting to sink in as to what to be looking for. I'm looking forward to applying this.

### *DAY 5 – Validity & Reliability (& Task #5)*

Although these topics seemed to be two of the more difficult ones addressed during the training, the teachers expressed that, not only did they learn more about validity and reliability, but also had a sense of how to apply the concepts to their assessments:

Learning more about reliability has been very helpful. I feel much more (or somewhat more) confident about writing more valid and reliable tests.

This class has definitely opened my eyes to test making. Many of these things I would never have considered. I know the more I critique my tests, the more effective they will be!

### *DAY 6 – Interpreting Standardized Test Data (& Task #6)*

This was undoubtedly one of the more highly anticipated topics for the teachers at the beginning of the workshop. With the more stringent expectations being placed on teachers and districts by No Child Left Behind (NCLB) and its associated Adequate Yearly Progress (AYP) requirements, these teachers really wanted to learn more about interpreting the standardized test scores of their students. They walked away from this day's session a bit overwhelmed, but much better informed:

Today was one of the most useful days of this workshop to me. I feel much more comfortable talking about and interpreting standardized tests and test results. I am looking forward to being able to look at this data more critically.

Oh, my – so much information . . . [but] Wow! I think I actually understand! The chart on the various ways to read a norm-referenced test was very helpful. Its one of those pages which will be removed from the class work and added to my planning book (my spot for all important info).

Today was a lot of information! The more we discussed it, the easier it was to interpret the scores.

We have to take [standardized tests] and they aren't going away anytime soon, so we should try and find a use for the data.

### *DAY 7 – Grading (& Task #7)*

Grading was a topic with which most of the teachers already felt comfortable. The session did, however, highlight several aspects of their individual grading systems on which the teachers began to reflect. One teacher considered the impact that performing statistical analyses, and ultimately making revisions to classroom tests, might have on her students' grades:

I do feel a bit more confident when I grade assignments, such as tests, but I am wondering how re-doing/improving my tests will be reflected in my grading system. I do need to make sure my grades are fair, accurate, and based on reliable, valid data.

Other teachers commented that grading items that are more subjective in nature (e.g. attendance, participation, effort) can have a substantial impact on a student's grades:

I never thought about the problem with how really subjective things can hurt a grade.

When we learned about grading systems, [it] opened my eyes to many issues that can contribute to a good or poor grading system.

[Today's] discussion on grading . . . showed me that others encounter the same problems. How to ensure that the grades assigned reflect the students' abilities and achievements.

How do I know which work to score and which to just give informative feedback? Grading is assumed to be a given that all teachers understand grading procedures, although it isn't always easy!

I believe that I will continue to assign regular homework but weight tests heavier. I may also have more quizzes so that tests aren't such a shock for kids to have to work on their own.

### *DAY 8 – Performance Assessment (& Task #8)*

Most of the interest, and comments, revolved around the idea of making quality rubrics to accompany performance assessment tasks. Most of the teachers reported that they had developed rubrics, but were unsure about the quality of their rubrics:

I do not know how to put together a good, solid rubric. I know how necessary it is for me to assess my students using a rubric, but I do not feel confident doing so. I want to learn how to construct a rubric that is accurate and clearly explains to my students exactly what is expected on each assessment.

This same teacher later commented:

I really have a much better understanding of performance assessments and constructing rubrics. I have always tried to work around using rubrics due to the fact that I just did not know how to make one. After today, I believe I can construct my own analytical rubrics.

Other teachers reiterated the initial concern about the quality of their rubrics:

I have [now] realized that the rubrics I have created in the past were too vague.

I'm excited about learning about performance assessments. I realized that I do these types of assessments, I just don't call them by their formal name. I make rubrics on my own, but I don't necessarily know if they're good or not.

Many teachers I work with have me create rubrics for them. They call me the 'Queen of Rubrics'. [However,] I know they are not set up as well as they should be. [The task] really made me think about how much needs to go into creating rubrics . . . !

After this day's lesson, nearly all of the teachers commented positively about now knowing how to improve the quality of their rubrics.

### *DAY 9 – Ethical Issues (& Task #9)*

Only one teacher provided comments in the journal about this day's topic of ethical issues related to classroom assessment. She commented that while she was very mindful

of unethical behaviors, she knew of many colleagues who exhibited questionable behaviors related to assessment:

I am very careful about the way I conduct my teaching as so nothing could be deemed 'questionable'. Many teachers want their students to do well, so they 'teach to the test'. This practice does not teach the students anything. Students need to learn test-taking skills and be exposed to the types of questions that will be on the test. Teachers should never use the test as a form of test preparation. I teach my students test-taking strategies that they could potentially use on any type of test.

### *Final Journal Entry – Overall Reflection*

Generally speaking, these seven teachers believed that the intensive, two-week workshop on classroom assessment was highly beneficial to their work as classroom teachers. Many of them expressed hesitancy at the beginning of the workshop, but felt much more confident following its completion. One teacher commented:

Super workshop! I was very hesitant when registering for this class because of the content and my lack of knowledge of assessment. I feel more comfortable about the process of assessment. This fall I will have a whole new outlook on how I will be assessing my students. This class allowed me to better understand the assessment process and all that it entails. I was glad to then be able to apply this information (in the tasks) to something I will now be able to use in my classroom. I am excited to share all that I learned in this workshop with the other teachers in my building.

All seven of the teachers commented on the overall impact that they believed the workshop had on them, focusing on its intensity and on its importance:

All I can say is WOW! . . . I have learned so much and looked at things in different ways! Overall, many questions I had about assessment were made much clearer.

I have to be totally honest. I was not at all looking forward to the class. I needed the credit but wasn't all that interested in the topic. But, after the second day, my interest was held. If you had said, what is the difference between a norm and a criterion I would have given you a blank look. I know what they look like, I've seen them!! The information about reading standardized test [scores] is alone worth much more than the cost of this course.

Wow! Where do I even begin? As I think back, it seems like it's been so much longer than two weeks. It's also amazing how much more confident I feel about assessment. I still have a long way to go before I can say I'm 100% assessment-literate, but I feel so much more knowledgeable about the topic. I feel like I could go back to school and share this info with other teachers and actually be able to explain it! This was a great experience and an eye opener too! I think in-depth assessment literacy classes should be mandatory for all inservice teachers.

A couple of the teachers believed that the most valuable aspect of the workshop was the fact that it 'forced' teachers to stop and critically examine what they do in their classes from an assessment perspective:

I think the biggest value of this workshop for me has been dedicating a two-week block of my life to intensely think about/reflect on the assessment of my students. I consider it to be of great value to be reminded of what we should keep in mind when assessing, or any other feature of teaching.

The time for reflection, or the dedicated time for reflection on this topic was invaluable. Plus it is a topic that pervades what I do. It applies to nearly my entire job! I need to be able to assess well to do my job well.

This class has been good for me because it has made me stop and think critically about my own assessment system. I wish I would have had this training earlier on, because I've never felt like assessment was my strong point.

One final comment from one of the participants may best sum up these sentiments shared by all of the teacher participants:

I feel like I have been in an assessment tornado and survived! This was the most intense workshop that I have ever been to – which is a great thing. I feel that walking into this workshop I know very little in comparison to what I know now. And the greatest thing is that now I actually know how to apply them to my own classroom.

## Conclusions

The day-to-day work of classroom teachers is multifaceted, to say the least. Arguably, none of these daily responsibilities is more central to the work of teachers than that of assessing student performance (Mertler, 2003b). Previous studies have reported that teachers feel – are actually *are* – unprepared to adequately assess their students (e.g. Mertler, 1998, 1999; Plake, 1993). They often believe that they have not received sufficient training in their undergraduate preparation programs, as well as in their graduate programs, in order to feel comfortable with their skills in conducting assessments and in making assessment decisions.

This two-week intensive workshop in classroom assessment proved to be very meaningful to the teachers involved. Their knowledge of assessment concepts and terminology, as measured by the *ALI*, improved dramatically as a result of the workshop. Further evidence of the impact of the assessment training was provided in the teachers' reflective journals. All of the teachers reflected on the training as a positive experience that made them feel more comfortable and confident with the process of assessing student academic performance. The teachers were very forthcoming about their limitations with respect to assessment prior to the workshop. Following the workshop, all seven indicated that they had a new perspective on assessing their students that they could not wait to take back to their classrooms in the fall, and perhaps even share with other teachers in their buildings. Comments provided in their reflective journals demonstrated the importance that this workshop had for the teachers involved.

Perhaps, performance-based inservice training sessions, which focus on *applied* assessment decision-making, could prove to be beneficial to a majority of classroom teachers. They may increase the relevance of assessment to inservice teachers, which may increase the likelihood that the assessment concepts addressed in the training will carry through to their classrooms and subsequent practices. Along those lines, the results of this study also have implications for the structure and evaluation of teacher professional development on other themes, and for applying this in-depth approach to professional development more broadly in order to meet other specific professional and educational requirements. The intensive approach, as well as the use of both

qualitative (in particular, the reflective journals) and quantitative evidence could, and should, be applied to other professional development.

Additionally, as voiced by teachers in this study, perhaps the most beneficial aspect of the training was its intensive nature. This permitted teachers to concentrate on and critically examine their own assessment knowledge and practices, in light of what they *should* know and *should* be doing. What remains to be seen, however, is the extent to which inservice training workshops, such as the one completed by the teachers in this study, facilitate a long-term – and, perhaps, lasting – impact on teachers' classroom assessment practice.

The author can be contacted via email at: mertler@bgsu.edu

## References

- American Federation of Teachers (AFT), National Council on Measurement in Education (NCME), & National Education Association (NEA) (1990) *Standards for Teacher Competence in the Educational Assessment of Students*. Online: <http://www.unl.edu/buros/bimm/html/article3.html> [accessed 29 March 2005].
- Brookhart, S. M. (2001) The *Standards* and classroom assessment research. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Dallas, TX, March (ERIC Document Reproduction Service No. 451 189).
- Campbell, C., Murphy, J. A. & Holt, J. K. (2002) Psychometric analysis of an assessment literacy instrument: applicability to preservice teachers. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH, October.
- Creswell, J. W. (2003) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 2nd edn. Thousand Oaks, CA: SAGE.
- Mertens, D. M. (2005) *Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative, and Mixed Methods*, 2nd edn. Thousand Oaks, CA: SAGE.
- Mertler, C. A. (1998) Classroom assessment practices of Ohio teachers. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL, October (ERIC Document Reproduction Service No. 428 085).
- Mertler, C. A. (1999) Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, 120, 285–96.
- Mertler, C. A. (2003a) Preservice versus inservice teachers' assessment literacy: Does classroom experience make a difference? Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH, October.
- Mertler, C. A. (2003b) *Classroom Assessment: A Practical Guide for Educators*. Los Angeles, CA: Pyrczak.
- Mertler, C. A. & Campbell, C. S. (2004) Assessing those who assess: Development of an instrument to measure teachers' assessment literacy. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH, October.
- Mertler, C. A. & Campbell, C. S. (2005) Measuring teachers' knowledge and application of classroom assessment concepts: development of the *Assessment Literacy Inventory*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada, April.
- O'Sullivan, R. S. & Johnson, R. L. (1993) Using performance assessments to measure teachers' competence in classroom assessment. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA, April.
- Plake, B. S. (1993) Teacher assessment literacy: teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(2), 21–7.
- Popham, W. J. (2003) Seeking redemption for our psychometric sins. *Educational Measurement: Issues and Practice*, 22(1), 45–8.
- Stiggins, R. J. (1995) Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238–45.
- Stiggins, R. J. (1999) Are you assessment literate? *High School Magazine*, 6(5), 20–3.
- Tashakkori, A. & Teddlie, C. (1998) *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Applied Social Research Methods Series, No. 46. Thousand Oaks, CA: SAGE.