

DOCUMENT RESUME

ED 482 277

TM 035 307

AUTHOR Mertler, Craig A.  
TITLE Preservice Versus Inservice Teachers' Assessment Literacy: Does Classroom Experience Make a Difference?  
PUB DATE 2003-10-00  
NOTE 29p.; Paper presented at the Annual Meeting of the Mid-Western Educational Research Association (Columbus, OH, October 15-18, 2003).  
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)  
EDRS PRICE EDRS Price MF01/PC02 Plus Postage.  
DESCRIPTORS \*Competence; \*Educational Assessment; \*Knowledge Level; Preservice Teacher Education; \*Preservice Teachers; \*Student Evaluation; \*Teachers

ABSTRACT

Assessing student performance is one of the most critical aspects of the job of a classroom teacher; however, many teachers do not feel adequately prepared to assess their students' performance. In order to measure and compare preservice and inservice teachers' "assessment literacy," both groups were surveyed using the Classroom Assessment Inventory (CALI), which was designed to parallel the "Standards for Teacher Competence in the Educational Assessment of Students." Inservice teachers performed highest on Standard 3—"Administering, Scoring, and Interpreting Results of Assessments"—and lowest on Standard 5—"Developing Valid Grading Procedures." Respondents were 67 preservice teachers and 197 inservice teachers. Preservice teachers performed highest on Standard 1—"Choosing Appropriate Assessment Methods"—and lowest on Standard 5—"Developing Valid Grading Procedures. Comparisons between the two groups revealed significant differences on five of seven competency areas, as well as on the total scores. In all cases where significant differences were found, the inservice teachers scored higher than their preservice counterparts. (Contains 2 tables and 25 references.) (Author/SLD)

PRESERVICE VERSUS INSERVICE TEACHERS'  
ASSESSMENT LITERACY:  
DOES CLASSROOM EXPERIENCE MAKE A  
DIFFERENCE?

PERMISSION TO REPRODUCE AND  
DISSEMINATE THIS MATERIAL HAS  
BEEN GRANTED BY

C. Mertler

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)

1

Craig A. Mertler, Ph.D.

([mertler@bgnet.bgsu.edu](mailto:mertler@bgnet.bgsu.edu))

Division of Educational Foundations & Inquiry  
School of Educational Leadership & Policy Studies  
College of Education & Human Development  
Bowling Green State University  
Bowling Green, OH 43403

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

This document has been reproduced as  
received from the person or organization  
originating it.

Minor changes have been made to  
improve reproduction quality.

Points of view or opinions stated in this  
document do not necessarily represent  
official OERI position or policy.

Paper presented at the annual meeting of the

Mid-Western Educational Research Association, Columbus, OH (Oct. 15–18, 2003)

Running head: PRESERVICE VERSUS INSERVICE TEACHERS' ASSESSMENT LITERACY...

BEST COPY AVAILABLE

### Abstract

Assessing student performance is one of the most critical aspects of the job of a classroom teacher; however, many teachers do not feel adequately prepared to assess their students' performance. In order to measure and compare preservice and inservice teachers' "assessment literacy," both groups were surveyed using the *Classroom Assessment Literacy Inventory (CALI)*, which was designed to parallel the *Standards for Teacher Competence in the Educational Assessment of Students*. Inservice teachers performed highest on Standard 3—*Administering, Scoring, and Interpreting the Results of Assessments* and lowest on Standard 5—*Developing Valid Grading Procedures*. Preservice teachers performed highest on Standard 1—*Choosing Appropriate Assessment Methods* and lowest on Standard 5—*Developing Valid Grading Procedures*. Comparisons between the two groups revealed significant differences on five of the seven competency areas, as well as on the total scores. In all cases where significant differences were found, the inservice teachers scored higher than their preservice counterparts.

PRESERVICE VERSUS INSERVICE TEACHERS' ASSESSMENT LITERACY:  
DOES CLASSROOM EXPERIENCE MAKE A DIFFERENCE?

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. For example, aspects of a teacher's job that are impacted by assessment include, but are not limited to, the following:

- guiding decisions about large-group instruction;
- developing individualized instructional programs;
- determining the extent to which instructional objectives have been met;
- providing information for administrative decisions, such as promotion, retention, or graduation; and
- providing data for state or federal programs.

With respect to classroom assessment, there exists a paradox in our educational system. Accurate assessment of achievement is being more urgently called for at the district, state, and national levels (Rogers, 1991). Various reform efforts are forcing teachers to be held accountable for their assessment of student learning. However, teachers do not feel adequately prepared to meet this challenge. Classroom teachers are calling for more training due to their perceived lack of preparedness to assess their students, citing weaknesses in their undergraduate preparation programs (Rogers, 1991).

## Review of Related Literature

Discussed in this review are specific works related to (1) general research on classroom assessment, (2) *The Standards for Teacher Competence in the Educational Assessment of Students*, (3) assessment literacy, and (4) specific research studies that have been conducted on *The Standards* and assessment literacy.

### *Research on Classroom Assessment*

It has been estimated that teachers spend up to 50 percent of their time on assessment-related activities (Plake, 1993). Regardless of the amount of time spent on it, classroom assessment is a vitally important teaching function; it contributes to every other teacher function (Brookhart, 1998, 1999b). Assessment is used for numerous purposes: to diagnose student needs, to group students, to grade students, to evaluate instruction, to motivate students, etc. (Stiggins, 1999a). Sound assessment and grading practices help teachers to improve their instruction, improve students' motivation to learn, and increase students' levels of achievement (Brookhart, 1999a). According to Stiggins (1999a), "The quality of instruction in any ... classroom turns on the quality of the assessments used there" (p. 20). For all of these reasons, the information resulting from classroom assessments must be meaningful and accurate; i.e., the information must be valid and reliable (Brookhart, 1999a).

In recent years, public and governmental attention has shifted to school achievement as evidenced by performance on standardized achievement tests (Campbell, Murphy, & Holt, 2002). Additionally, there has been an increase in expectations regarding teachers' assessment expertise. Teachers have been required to develop classroom assessments that align curriculum with state standards as a means of improving

test scores (Campbell, Murphy, & Holt, 2002). New research on the relationship between classroom assessments and student performance on standardized tests reveals that improving the quality of classroom assessments can increase average scores on large-scale assessments as much as  $3/4$  of a SD (as much as 4 grade equivalents or 15-20 percentile points), representing a huge potential (Stiggins, 1999a). This is important research since it makes a connection between the quality of assessment in the classroom and assessment resulting from standardized testing programs.

Ironically, in this age of increase in emphasis on testing and assessment, many colleges of education and state education agencies do not require preservice teachers to complete specific coursework in classroom assessment (Campbell, Murphy, & Holt, 2002; O'Sullivan & Johnson, 1993). This continues to be an interesting phenomenon since many inservice teachers report that they are not well prepared to assess student learning (Plake, 1993). Furthermore, these teachers often claim that the lack of adequate preparation is largely due to inadequate preservice training in the area of educational measurement (Plake, 1993). Brookhart (2001) also cites literature that calls for an increase in emphasis in teacher preparation programs on classroom assessment and a decrease in emphasis on large-scale testing. Studies have generally concluded that teachers' skills in both areas are limited.

Three methods used have been used to investigate teachers' assessment practices, as well as their levels of preparation to assess students: surveys of attitudes, beliefs, and practices; tests of assessment knowledge; and reviews of teachers' actual assessments (Brookhart, 2001). Regardless of the method used, research has documented that teachers' assessment skills are generally weak (Campbell, Murphy, & Holt, 2002).

Stiggins (2001) is in agreement when he states that we are seeing unacceptably low levels of assessment literacy among practicing teachers and administrators in our schools. He continues by stating that this assessment *il*-literacy has resulted in inaccurate assessment of students, causing them to fail to reach their full potential.

With respect to teachers' assessment practices, for example, Mertler (1999) found that teachers did not perform statistical analyses of test data (e.g., estimating reliability, conducting item analyses) very often. Furthermore, teachers indicated that they followed specific steps to insure validity and reliability about half of the time or less (Mertler, 2000). When asked to list specific steps that teachers follow to insure *validity*, a wide variety ( $N = 611$ ) of responses were offered by the teachers. Only half of those responses provided procedures that were appropriate (or at least approximate); about one-third were simply not appropriate (e.g., "I check reliability," "I use statistical analyses," etc.); less than 20% focused on content-related evidence of validity (which is most appropriate for teacher-made tests); numerous teachers provided "procedures" that were troubling, to say the least (e.g., "It can't be done," "I don't have time," "I don't know what validity even is," "teachers don't have time for this," and "You'll just figure out what works for you").

When asked to list specific steps that teachers follow to insure reliability, again a wide variety ( $N = 431$ ) of responses offered (Mertler, 2000). Only 10% indicated that they used statistical analyses (the appropriate response); over half said they are automatically reliable if you use teacher-made tests, or provided other troubling comments (e.g., "There are no specific steps," "I have no time to do this," "Is there really a difference between validity and reliability?," and "Worrying about reliability is way down on list of priorities").

With respect to teachers' levels of assessment preparation, for example, over 70% 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 6 years. Those who had had previous coursework/training scored significantly higher on a test of assessment literacy than those who hadn't, but the difference was less than one point (Plake, 1993).

When inservice teachers in a statewide study were asked about the level of preparedness to assess student learning resulting from their teacher preparation programs, the median response was "slightly prepared" (Mertler, 1999). When asked about their current level of preparedness, the median response improved to "somewhat prepared." Mertler (1999) concluded that this potentially implies that teachers tend to develop assessments skills on the job, as opposed to structured environments such as courses or workshops. Stiggins (1999a) has reiterated this implication, stating that many teachers are left unprepared to assess student learning as a result of both preservice and graduate training; they acquire what assessment "expertise" and skills they possess while on the job.

Brookhart (2001) has quite accurately summarized the research on teachers' assessment practices when she states that teachers apparently do better at classroom applications than at interpreting standardized tests (likely due to nature of their work). Additionally, they lack expertise at test construction, and they do not always use valid grading procedures.



*“The Standards for Teacher Competence in the Educational Assessment of Students”*

*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. This joint effort began in 1987 in order to “develop standards for teacher competence in student assessment out of concern that the potential educational benefits of student assessments be fully realized” (AFT, NCME, & NEA, 1990). They were originally developed in order to address the problem of inadequate assessment training for teachers (AFT, NCME, & NEA, 1990).

According to *The Standards* (AFT, NCME, & NEA, 1990), *assessment* is defined as “the process of obtaining information that is used to make educational decisions about students, to give feedback to the student about his or her progress, strengths, and weaknesses, to judge instructional effectiveness and curricular adequacy, and to inform policy.” *The Standards*, of which there are seven, provide criteria for teacher competence with respect to the various components of this definition of assessment. *The Standards for Teacher Competence in the Educational Assessment of Students* consists of the following seven principles:

1. *Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.*
2. *Teachers should be skilled in developing assessment methods appropriate for instructional decisions.*
3. *The teacher should be skilled in administering, scoring and interpreting the results of both externally produced and teacher-produced assessment methods.*

4. *Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.*
5. *Teachers should be skilled in developing valid pupil grading procedures that use pupil assessments.*
6. *Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.*
7. *Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.*

The *Standards* acknowledge and specify the importance of teacher education and professional development in the area of classroom assessment (Brookhart, 2001). All 7 standards apply to teachers' development and use of classroom assessments of instructional goals and objectives that form basis for classroom instruction. Standards 3, 4, 6, 7 also apply to large-scale assessment, including administering, interpreting, and communicating assessment results, using information for decision making, and recognizing unethical practices (Brookhart, 2001).

*What is "Assessment Literacy"?*

Several times in this paper, the term "assessment literacy" has been mentioned. Since *The Standards* and the concept of assessment literacy are central to this study, it is imperative that the term be defined or otherwise described here. Assessment literacy has been defined as "the possession of knowledge about the basic principles of sound assessment practice, including terminology, the development and use of assessment methodologies and techniques, familiarity with standards of quality in assessment...and

familiarity with alternative to traditional measurements of learning” (Paterno, 2001). An alternative, simpler definition is offered by the North Central Regional Educational Laboratory: “the readiness of an educator to design, implement, and discuss assessment strategies” (n.d.).

Others have chosen not to formally define assessment literacy, but rather to describe the characteristics of those who possess it. One such characterization is as follows:

Assessment literate educators recognize sound assessment, evaluation, communication practices; they

- understand which assessment methods to use to gather dependable information and student achievement.
- communicate assessment results effectively, whether using report card grades, test scores, portfolios, or conferences.
- can use assessment to maximize student motivation and learning by involving students as full partners in assessment, record keeping, and communication (Center for School Improvement and Policy Studies, Boise State University, n.d.).

Another similar description is provided by Stiggins (1995), who states that “Assessment literates 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 continues by stating 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).

*Research on Assessment Literacy and "The Standards"*

Numerous research studies have been conducted over the past 10 years that have addressed one or more of the seven *Standards* (Brookhart, 2001). However, only one (Plake, 1993) has addressed *all* teacher competencies—as specified by *The Standards*—for inservice teachers. Additionally, one other study (Campbell, Murphy, & Holt, 2002) has attempted to apply *The Standards* to groups of undergraduate preservice teachers. Finally, one other study attempted to integrate *The Standards* into a graduate-level course through the use of performance assessments (O'Sullivan & Johnson, 1993).

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 used in the study. The survey instrument (the *Teacher Assessment Literacy Questionnaire*) consisted of 35 items (5 per standard). Items were developed as application-type questions—realistic and meaningful to teachers' actual practices. The instrument went through extensive content validation and pilot testing. A representative sample from around country was selected to participate; a total of 98 districts in 45 states participated, with a total usable sample of 555 surveys (Plake, 1993). The KR-20 reliability for the entire test was equal to .54 (Plake, Impara, & Fager, 1993).

Teachers answered an average of slightly more than 23 out of 35 items correct. The teachers' highest performance occurred on Standard 3—*Administering, Scoring, and Interpreting the Results of Assessments* ( $M = 3.96/5.00$ ); the lowest performance occurred on Standard 6—*Communicating Assessment Results* ( $M = 2.70/5.00$ ). On 10 of the 35 items, 90% or more of teachers answered the item correctly. These items addressed issues including selecting appropriate assessments, acceptable test taking behavior for

standardized testing situations, explanation of the basis for a grade to a child's parent, and the recognition of unethical practices in standardized test administration. On 5 items, less than 30% answered correctly. Two of the five came from Standard 5—*Developing Valid Grading Procedures*. Only 13% answered correctly an item that focused on steps to increase reliability of a test score. The two remaining items with low performance addressed Standard 7—*Recognizing Unethical or Illegal Practices*).

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 (ALI)* was administered to 220 undergraduate students following course in tests and measurement. The course included topics such as creating and critiquing various methods of assessment, discussing ethical considerations related to assessment, interpreting and communicating both classroom and standardized assessment results, and discussing and evaluating psychometric qualities (i.e., validity and reliability) of assessments.

The data from the undergraduate preservice teachers exhibited a higher level of reliability ( $\alpha = .74$ ) than their inservice counterparts in the Plake et al. study (Campbell, Murphy, & Holt, 2002). The preservice teachers ( $M = 21$ ) averaged two fewer questions answered correctly than did the inservice teachers ( $M = 23$ ). Six items (numbers 5, 7, 22, 28, 31, and 35) demonstrated poor item discrimination values ( $< .20$ ). The inservice teachers in the Plake et al. study scored higher than the preservice teachers on all but one standard (Standard 1—*Choosing Appropriate Assessment Methods*). The preservice teachers scored highest on Standard 1, whereas the inservice teachers scored highest on

Standard 3. Both groups of teachers scored lowest on Standard 6—*Communicating Assessment Results*.

Finally, a third study attempted to integrate *The Standards* into a graduate-level course in measurement. O'Sullivan and Johnson (1993) designed a course that incorporated performance assessments ( $N = 8$ ) which were aligned with *The Standards*. Teachers were pretested ( $M = 24.2$ ) and posttested ( $M = 27.3$ ) using the Plake et al. instrument. The results indicated a slight improvement in assessment literacy over the duration of the course.

#### Purpose of the Study

It was the intent of this study to investigate the concept of “assessment literacy” and attempt to measure it as delineated by *The Standards for Teacher Competence in the Educational Assessment of Students*. Specifically, the purposes of this study were: (1) to measure and describe the relative levels of assessment literacy for both preservice and inservice teachers, and (2) to statistically compare the relative levels of assessment literacy for these two groups. This is the first study that attempts to measure assessment literacy for both preservice and inservice teachers and statistically compare the results.

The research questions addressed in the study were:

Research Question 1: What is the level of assessment literacy, as measured by the *Classroom Assessment Literacy Inventory*, for preservice teachers?

Research Question 2: What is the level of assessment literacy, as measured by the *Classroom Assessment Literacy Inventory*, for inservice teachers?

Research Question 3: How does the assessment literacy of preservice teachers compare to that of inservice teachers? Are there any significant differences between the two groups?

## Methods

### *Participants*

During the fall of 2002, the researcher surveyed both preservice and inservice teachers with respect to their assessment literacy. The group of preservice teachers was comprised of 67 undergraduate students, all majoring in secondary education, at a midwestern university. At the time of data collection, they were enrolled in methods courses (i.e., the term preceding student teaching) and had just completed a course in classroom assessment. The group of inservice teachers consisted of 197 teachers, representing nearly every district and school in a three-county area surrounding the same institution. The schools were selected based on convenience due to their geographic location. All grade levels and content areas were represented in the final sample.

### *Instrumentation*

Both groups of teachers were surveyed using an instrument titled the *Classroom Assessment Literacy Inventory*, or *CALI*, which was adapted from a similar instrument called the *Teacher Assessment Literacy Questionnaire* (Plake, 1993; Plake, Impara, & Fager, 1993). This inventory is based on the *Standards for Teacher Competence in the Educational Assessment of Students* (AFT, NCME, & NEA, 1990). The *CALI* consisted of the same 35 content-based items (five per standard) with a limited amount of rewording (e.g., changing some names of fictitious teachers, changing word choice to improve clarity, etc.), as well as 7 demographic items.

The original instrument has been shown to have reasonable reliability with both inservice teachers,  $r_{KR-20} = .54$  (Plake, Impara, & Fager, 1993), and preservice teachers,  $\alpha = .74$  (Campbell, Murphy, & Holt, 2002). Furthermore, the original instrument was subjected to a thorough content validation, including reviews by members of the National Council on Measurement in Education and a pilot study with and feedback from practicing teachers and administrators.

### *Procedures*

Inservice teachers were sent the *CALI* in both paper and Web-based formats. Two weeks after the initial mailing of the paper version and posting of the Web-based version, teachers were sent a reminder about completing the instrument. The instrument was administered to the preservice teachers at the final class meeting in their classroom assessment course. They were informed that their individual decision about participation, as well as their individual score on the instrument, would in no way affect the grade received for the course.

### *Analyses*

Descriptive analyses at the individual item level included frequencies and reliability analyses; descriptive analyses were also conducted for the seven composite scores (i.e., based on *The Standards*). Inferential analyses included *t*-test comparisons (evaluated at an  $\alpha$ -level equal to .05) of the preservice to inservice teacher mean scores for each of seven composite scores, as well as the total score for the entire instrument. All analyses were conducted using SPSS (v. 11).



## Results

The results that follow are presented by each individual research question.

**Research Question 1: What is the level of assessment literacy, as measured by the *Classroom Assessment Literacy Inventory*, for preservice teachers?**

The data resulting from the preservice teacher group ( $N = 67$ ) demonstrated a reasonably good level of internal consistency reliability,  $\alpha = .74$ . On average, preservice teachers answered slightly less than 19 out of 35 items correctly. Out of the seven competency areas, as delineated by *The Standards*, the highest overall performance for preservice teachers was found for Standard 1—*Choosing Appropriate Assessment Methods* ( $M = 3.25$ ; maximum possible score = 5). The lowest performance was found for Standard 5—*Developing Valid Grading Procedure* ( $M = 2.06$ ). The results for the preservice teachers for each of the seven standards are presented in Table 1.

-----  
 Insert Table 1 here  
 -----

On only 4 of the 35 items did 90% or more of the preservice teachers answer the item correctly. One item each came from Standard 1—*Choosing Appropriate Assessment Methods* and Standard 2—*Developing Appropriate Assessment Methods*; two items came from Standard 3—*Administering, Scoring, and Interpreting the Results of Assessments*.

On five of the 35 items, 25% or fewer answer the item correctly. One item came from Standard 2—*Developing Appropriate Assessment Methods*; two items each came from Standard 5—*Developing Valid Grading Procedures* and Standard 7—*Recognizing Unethical or Illegal Practices*.

**Research Question 2: What is the level of assessment literacy, as measured by the *Classroom Assessment Literacy Inventory*, for inservice teachers?**

The data resulting from the inservice teacher group ( $N = 197$ ) demonstrated a mediocre level of internal consistency reliability,  $\alpha = .57$ . On average, inservice teachers answered slightly less than 22 out of 35 items correctly. Out of the seven competency areas, the highest overall performance for inservice teachers was found for Standard 3—*Administering, Scoring, and Interpreting the Results of Assessments* ( $M = 3.95$ ; maximum possible score = 5). The lowest performance was found for Standard 5—*Developing Valid Grading Procedures* ( $M = 2.06$ ). The results for the inservice teachers for each of the seven standards are presented in Table 2.

-----  
 Insert Table 2 here  
 -----

On 8 of the 35 items, 90% or more of the inservice teachers answered the item correctly. Two items each came from Standard 1—*Choosing Appropriate Assessment Methods*, Standard 2—*Developing Appropriate Assessment Methods*, Standard 3—*Administering, Scoring, and Interpreting the Results of Assessments*, and Standard 7—*Recognizing Unethical or Illegal Practices*.

On six of the 35 items, 25% or fewer answered the item correctly. One item came from Standard 2—*Developing Appropriate Assessment Methods*; three items came from Standard 5—*Developing Valid Grading Procedures*; and two items came from Standard 7—*Recognizing Unethical or Illegal Practices*.

**Research Question 3: How does the assessment literacy of preservice teachers compare to that of inservice teachers? Are there any significant differences between the two groups?**

Standard and total scores for the two groups of teachers were compared by conducting independent-samples *t*-tests ( $\alpha = .05$ ). Examination of the results revealed that significant differences existed between the two groups for scores on 5 of the 7 Standards, as well as for the total scores. In all cases where there were significant differences, the inservice teachers scored significantly higher (i.e., they were more assessment literate) than their preservice counterparts. The largest discrepancies were found for Standard 3, the total score, and Standard 4, respectively. For Standard 3, the inservice teachers scored significantly higher ( $M = 3.95, SD = .95$ ) than the preservice teachers ( $M = 3.24, SD = 1.00$ ),  $t(262) = 5.23, p < .05$ , two-tailed. For the total score, the inservice teachers scored significantly higher ( $M = 21.96, SD = 3.44$ ) than the preservice teachers ( $M = 18.96, SD = 4.65$ ),  $t(262) = 4.85, p < .05$ , two-tailed. For Standard 4, once again the inservice teachers scored significantly higher ( $M = 3.36, SD = 1.08$ ) than the preservice teachers ( $M = 2.67, SD = 1.19$ ),  $t(262) = 4.36, p < .05$ , two-tailed. Significant differences were also found for Standards 1, 2, and 7. There were no significant differences found between the groups for Standards 5 and 6. Interestingly, both groups performed the poorest—and at the same exact level—on Standard 5. The results of all *t*-tests are presented in Table 3.

-----  
 Insert Table 3 here  
 -----

## Discussion

Many of the results of this study parallel those of an earlier study (Plake, 1993; Plake, Impara, & Fager, 1993) that used the original version of the instrument and focused on the assessment literacy of inservice teachers. With respect to overall performance on the 35 items, the average score was equal to 22 items answered correctly—quite similar to the average score of 23 obtained by Plake (1993). In the earlier study, the highest mean performance for a given competency area was on Standard 3—*Administering, Scoring, and Interpreting the Results of Assessments*; the lowest performance was on Standard 6—*Communicating Assessment Results*. In the present study, the highest mean performance was also on Standard 3; the lowest was on Standard 5—*Developing Valid Grading Procedures*. Reliability analyses also revealed similar values for internal consistency ( $\alpha = .54$  and  $.57$  for the original study and the study at hand, respectively).

The results for the preservice teachers also reflected those from a recent study, which also used the original instrument but collected data from preservice teachers (Campbell, Murphy, & Holt, 2002). In that study, the highest mean performance was on Standard 1—*Choosing Appropriate Assessment Methods*; the lowest performance was on Standard 6—*Communicating Assessment Results*. In the present study, the highest mean performance was also on Standard 1; the lowest was on Standard 5—*Developing Valid Grading Procedures*. Reliability analyses revealed identical values for internal consistency ( $\alpha = .74$  for both the original study and the study at hand).

Comparisons between preservice and inservice teachers of the seven competency area scores revealed significant differences on five of the seven areas, as well as on the

total scores. In all cases where significant differences were found, the inservice teachers scored higher than their preservice counterparts. Both groups demonstrated their poorest performance on Standard 5—*Developing Valid Grading Procedures*, followed closely by Standard 6—*Communicating Assessment Results*.

Research has shown that traditional teacher preparation courses in classroom assessment are not well matched with what teachers need to know for classroom practice (Schafer, 1993). The traditional focus has historically been on large-scale (standardized) testing (Schafer, 1993), although this trend is changing. One course in assessment and measurement may truly be insufficient to cover everything teachers need to know.

This fact is made even more troublesome when considering that many teacher preparation institutions and states do not even require a course in assessment (Campbell, Murphy, & Holt, 2002; Shafer, 1993). As of January 1998, only 15 states had teacher certification standards that required competence in assessment, and 10 states explicitly required a course in assessment; however, 25 states held no expectation of competence in assessment (Stiggins, 1999b). The majority of states and institutions simply embed assessment content into other teacher education coursework; students then learn about assessment and measurement from instructors who typically possess no expertise in educational assessment (Quilter, 1999).

However, instruction from individuals with expertise in educational assessment may not be enough. It may be more important, not that the instruction is presented by experts, but that these measurement specialists better understand the reality of K–12 classrooms. Specifically, it is important that they understand that assessment is an integral component of instruction and goals for student learning (McMillan, 2001;

Pilcher, 2001). Teachers have indicated that they are more concerned with the day-to-day issues related to the application of assessment processes and less with fundamental measurement principles (Rogers, 1991). Hopefully, then, those who teach courses in assessment and measurement can teach preservice teachers to see this vital connection between assessment and instruction, making assessment more applicable to their views of teaching.

With respect to the concept of assessment literacy, Popham (2003) has called for an increased effort among the measurement community at large to promote assessment literacy on the part of policymakers, practitioners (including teachers, administrators, and counselors), public, and parents. A more assessment literate citizenry is less likely to tolerate misuse of assessment and, specifically, assessment results. Stiggins (1995) offers several guiding principles for educators to follow in order to promote assessment literacy. These guiding principles suggest that educators should:

- start with a clear purpose for assessment,
- focus on achievement targets,
- select appropriate assessment methods,
- adequately sample student achievement, and
- avoid bias and distortion.

Stiggins (1995) continues by stating that these standards of assessment quality are not negotiable, nor is the expectation that they be met every time educators assess student achievement. However, research shows that these standards are seldom met—due to fear of assessment and evaluation, insufficient time to assess properly, or public perceptions of assessment practices.

The day-to-day work of classroom teachers is multifaceted, to say the least. However, none of these daily responsibilities is more important—or more central—to the work of teachers than that of assessing student performance (Mertler, 2003). Previous studies have reported that teachers feel—and actually *are*—unprepared to adequately assess their students (e.g., Mertler, 1999, 1998; Plake, 1993). They often believe that they have not received sufficient training in their undergraduate preparation programs in order to feel comfortable with their skills in making assessment decisions. This, coupled with the fact that inservice teachers outscored preservice teachers on nearly every subscale in this study, may raise substantial questions about the usefulness—or, perhaps more importantly, the appropriateness—of assessment training in preservice teacher education programs.

Another question worthy of consideration—and further research—is whether or not a majority of assessment training is an “on-the-job” type of training; in other words, are assessment skills best learned through *classroom experience* as a teacher, perhaps once teachers can place the notion of “assessment” in a specific context, as opposed to learning them as an undergraduate? Does undergraduate training provide the necessary foundation for this on-the-job training? At a minimum, the present study highlights specific competency areas—namely, developing valid grading procedures and communicating assessment results—where both preservice and inservice teachers need remediation and additional support. Additionally, the measurement community must take on the responsibility of improving assessment literacy among all educational stakeholders.

## References

- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *The Standards for Competence in the Educational Assessment of Students*. Retrieved July 22, 2003, from <http://www.unl.edu/buros/article3.html>
- Brookhart, S. M. (1998). *Teaching about grading and communicating assessment results*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA. (ERIC Document Reproduction Service No. 419838)
- Brookhart, S. M. (1999a). *The art and science of classroom assessment: The missing part of pedagogy*. Washington, D.C.: ERIC Clearinghouse on Higher Education and Office of Educational Research and Improvement.
- Brookhart, S. M. (1999b). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice*, 18(1), 5-13.
- 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. (ERIC Document Reproduction Service No. ED451189)
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002, October). *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.
- Center for School Improvement and Policy Studies, Boise State University. (n.d.). What is assessment literacy? In *Assessment literacy*. Retrieved July 24, 2003, from <http://csi.boisestate.edu/al/>



McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices.

*Educational Measurement: Issues and Practice*, 20(1), 20-32.

Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, 120(2), 285-296.

Mertler, C. A. (2000). Teacher-centered fallacies of classroom assessment validity and reliability. *Mid-Western Educational Researcher*, 13(4), 29-35.

Mertler, C. A. (2003). *Classroom assessment: A practical guide for educators*. Los Angeles, CA: Pyczak.

North Central Regional Educational Laboratory. (n.d.). *Indicator: Assessment*. Retrieved July 24, 2003, from <http://www.ncrel.org/engage/framework/pro/literacy/prolitin.htm>

O'Sullivan, R. G. & 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. (ERIC Document Reproduction Service No. 358156)

Paterno, J. (2001). Measuring success: A glossary of assessment terms. In *Building cathedrals: Compassion for the 21st century*. Retrieved July 24, 2003, from <http://www.angelfire.com/wa2/buildingcathedrals/measuringsuccess.html>

Pilcher, J. K. (2001). *The Standards and integrating instructional and assessment practices*. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Dallas, TX. (ERIC Document Reproduction Service No. ED451190)

Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 21-27.

- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-12, 39.
- Popham, W. J. (2003). Seeking redemption for our psychometric sins. *Educational Measurement: Issues and Practice*, 22(1), 45-48.
- Quilter, S. M. (1999). Assessment literacy for teachers: Making a case for the study of test validity. *Teacher Educator*, 34(4), 235-243.
- Rogers, W. T. (1991). Educational assessment in Canada: Evolution or extinction? *The Alberta Journal of Educational Research*, 37(2), 179-192.
- Schafer, W. D. (1993). Assessment literacy for teachers. *Theory Into Practice*, 32(2), 118-126.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238-245.
- Stiggins, R. J. (1999a). Are you assessment literate? *The High School Journal*, 6(5), 20-23.
- Stiggins, R. J. (1999b). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 18(1), 23-27.
- Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5-15.

Table 1

*Means and Standard Deviations for Preservice Teachers by Standard and Total Scores on Classroom Assessment Literacy Inventory*

Standard	Mean	Standard Deviation
Standard 1 <i>Choosing Appropriate Assessment Methods</i>	3.25	1.03
Standard 2 <i>Developing Appropriate Assessment Methods</i>	2.78	.83
Standard 3 <i>Administering, Scoring, and Interpreting the Results of Assessments</i>	3.24	1.00
Standard 4 <i>Using Assessment Results to Make Decisions</i>	2.67	1.20
Standard 5 <i>Developing Valid Grading Procedures</i>	2.06	.95
Standard 6 <i>Communicating Assessment Results</i>	2.27	1.32
Standard 7 <i>Recognizing Unethical or Illegal Practices</i>	2.69	1.13
-----		
Total Score	18.96	4.65

Note:  $N = 67$

Table 2

*Means and Standard Deviations for Inservice Teachers by Standard and Total Scores on Classroom Assessment Literacy Inventory*

Standard	Mean	Standard Deviation
Standard 1 <i>Choosing Appropriate Assessment Methods</i>	3.74	.86
Standard 2 <i>Developing Appropriate Assessment Methods</i>	3.18	.89
Standard 3 <i>Administering, Scoring, and Interpreting the Results of Assessments</i>	3.95	.95
Standard 4 <i>Using Assessment Results to Make Decisions</i>	3.36	1.08
Standard 5 <i>Developing Valid Grading Procedures</i>	2.06	.85
Standard 6 <i>Communicating Assessment Results</i>	2.57	1.23
Standard 7 <i>Recognizing Unethical or Illegal Practices</i>	3.10	.81
-----		
Total Score	21.96	3.44

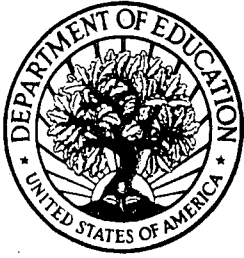
Note:  $N = 197$

Table 3

*t-Test Results for Comparisons of Scores for Preservice and Inservice Teachers*

Standard	Group	Mean	<i>t</i> -statistic	<i>p</i> -value
Standard 1	Preservice	3.25	3.79*	<.001
	Inservice	3.74		
Standard 2	Preservice	2.78	3.28*	.001
	Inservice	3.18		
Standard 3	Preservice	3.24	5.23*	<.001
	Inservice	3.95		
Standard 4	Preservice	2.67	4.36*	<.001
	Inservice	3.36		
Standard 5	Preservice	2.06	-.03	.975
	Inservice	2.06		
Standard 6	Preservice	2.27	1.69	.093
	Inservice	2.57		
Standard 7	Preservice	2.69	2.77*	.007
	Inservice	3.10		
Total Score	Preservice	18.96	4.85*	<.001
	Inservice	21.96		

\*  $p < .01$ .



**U.S. Department of Education**  
Office of Educational Research and Improvement (OERI)  
National Library of Education (NLE)  
Educational Resources Information Center (ERIC)



# REPRODUCTION RELEASE

(Specific Document)

**TM035307**

**I. DOCUMENT IDENTIFICATION:**

Title: <i>Preservice Versus Inservice Teachers' Assessment Literacy: Does Classroom Experience Make A Difference?</i>	
Author(s): <i>Craig A. Mertler</i>	
Corporate Source:	Publication Date: <i>Oct. 2003</i>

**II. REPRODUCTION RELEASE:**

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

The sample sticker shown below will be affixed to all Level 2A documents

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

*Sample*

---

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

*Sample*

---

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

*Sample*

---

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 1

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

Level 2A

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

Level 2B

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.  
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

*I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.*

Sign here, →  
please

Signature: <i>Craig A. Mertler</i>	Printed Name/Position/Title: <i>Craig A. Mertler, Assoc. Prof.</i>	
Organization/Address: <i>Div. of EDPI BGSU Bowling Green, OH 43403</i>	Telephone: <i>419-372-9357</i>	FAX: <i>419-372-9265</i>
	E-Mail Address: <i>mertler@bgsu.edu</i>	Date: <i>10-23-03</i>



(Over)

### III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

### IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

### V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse: <b>University of Maryland ERIC Clearinghouse on Assessment and Evaluation 1129 Shriver Lab, Bldg 075 College Park, MD 20742 Attn: Acquisitions</b>
--

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

**University of Maryland  
ERIC Clearinghouse on Assessment and Evaluation  
1129 Shriver Lab, Bldg 075  
College Park, MD 20742  
Attn: Acquisitions**