Mid-Western Educational Researcher

Volume 18 | Issue 4 Article 5

2005

The Role of Classroom Experience in Preservice and Inservice Teachers' Assessment Literacy

Craig A. Mertler
Bowling Green State University

Follow this and additional works at: https://scholarworks.bgsu.edu/mwer
How does access to this work benefit you? Let us know!

Recommended Citation
Mertler, Craig A. (2005) "The Role of Classroom Experience in Preservice and Inservice Teachers' Assessment Literacy," *Mid-Western Educational Researcher*: Vol. 18: Iss. 4, Article 5.
Available at: https://scholarworks.bgsu.edu/mwer/vol18/iss4/5

This Featured Article is brought to you for free and open access by the Journals at ScholarWorks@BGSU. It has been accepted for inclusion in Mid-Western Educational Researcher by an authorized editor of ScholarWorks@BGSU.

The Role of Classroom Experience in Preservice and Inservice Teachers' Assessment Literacy[1]

Craig A. Mertler
Bowling Green State University

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," two 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.*

Background

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). 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; U.S. Department of Education, n.d.). 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 increasing 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 reported that they are not well prepared to assess student learning (Plake, 1993). Furthermore, these teachers claimed that the lack of adequate preparation is largely due to inadequate preservice training in the area of educational measurement (Plake, 1993). Brookhart (2001) also cited 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.

Brookhart (2001) quite accurately summarized the research on teachers' assessment practices when she stated that teachers apparently do better at classroom applications than at interpreting standardized tests, perhaps due to the nature of their work.

What is "Assessment Literacy"?

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 who suggested assessment literacy is "the readiness

[1] This article was accepted for publication by the previous editorial team.

of an educator to design, implement, and discuss assessment strategies" (n.d.).

Others choose not to formally define assessment literacy, but rather to describe the characteristics of those who possess it. One such characterization proposes:

Assessment literate educators recognize sound assessment, evaluation, [and] 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.).

Stiggins (1995) provided another similar description when he stated 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 continued 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 or 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 for Teacher Competence in the Educational Assessment of Students"

The concept of assessment literacy is a key component of *The Standards for Teacher Competence in the Educational Assessment of Students* (AFT, NCME, & NEA, 1990). Additionally, *The Standards* are central to the study at hand, so it is imperative that they be described here. *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 of 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:

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

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

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

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

Standard 5—*Teachers should be skilled in developing valid pupil grading procedures that use pupil assessments*.

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

Standard 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).

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) addressed *all* teacher competencies—as specified by *The Standards*—for inservice teachers. Additionally, one other study (Campbell, Murphy, & Holt, 2002) attempted to apply *The Standards* to groups of undergraduate preservice teachers.

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 in-

strument (the *Teacher Assessment Literacy Questionnaire*) consisted of 35 items (5 per standard). Items were developed as application-type questions that were realistic and meaningful to teachers' actual practices. The instrument went through extensive content validation and pilot testing. A representative sample from around the country was selected to participate and a total of 98 districts in 45 states participated. There was a total usable sample of 555 surveys (Plake, 1993) and 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 a 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 (α = .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*.

Purpose of the Study

My intent in this study was 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 specific research questions addressed in the study were:

- How does the assessment literacy of preservice teachers compare to the assessment literacy 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 scheduled during 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. The researcher assigned pseudonyms to represent the names of the teachers and changed word choice to improve clarity. Additionally, 7 demographic items were included. The items were grouped by Standard; Table 1 shows the alignment of items with their respective Standard.

The 35 items presented the respondents with assessment-related scenarios, followed by a question with a specific

correct answer. Each item had the same format featuring four options presented in a multiple choice format with one option being the correct response. The complete instrument can be viewed at the following URL: http://edhd.bgsu.edu/mertler/cali.html.

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.

Table 1
Alignment of The Standards with Respective CALI Items

Standard	Item Numbers
Standard 1 *Choosing Appropriate Assessment Methods*	#1, 2, 3, 4, 5
Standard 2 *Developing Appropriate Assessment Methods*	#6, 7, 8, 9, 10
Standard 3 *Administering, Scoring, and Interpreting the Results of Assessments*	#11, 12, 13, 14, 15
Standard 4 *Using Assessment Results to Make Decisions*	#16, 17, 18, 19, 20
Standard 5 *Developing Valid Grading Procedures*	#21, 22, 23, 24, 25
Standard 6 *Communicating Assessment Results*	#26, 27, 28, 29, 30
Standard 7 *Recognizing Unethical or Illegal Practices*	#31, 32, 33, 34, 35

Note: The *Classroom Assessment Literacy Inventory* (*CALI*) can be viewed at http://edhd.bgsu.edu/mertler/cali.html

Table 2
Demographic Characteristics of Inservice and Preservice Teachers Responding to the CALI

Demographic Characteristic	Level	Inservice Teachers ($n = 197$)	Preservice Teachers ($n = 67$)
Gender	Female	77%	43%
	Male	21%	57%
Teaching Level	Elementary	57%	—[a]
	Secondary	26%	—[a]
Education Level	Pre-Bachelors	0%	100%
	Bachelors	29%	0%
	Masters	67%	0%
Years of Experience	None	0%	100%
	1-5	16%	0%
	6-10	14%	0%
	11-15	17%	0%
	16-20	12%	0%
	21-25	22%	0%
	> 25	18%	0%

[a] Preservice teachers could not provide responses to the demographic item addressing teaching level.

Analyses

Descriptive analyses at the individual item level included frequencies and reliability analyses. Descriptive analyses were also conducted for the seven composite scores based on *The Standards*. Inferential analyses included *t*-test comparisons, evaluated at an α-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

One-hundred ninety-seven (N = 197) inservice teachers completed the instrument. Seventy-seven percent of the sample was female; 21% was male. With respect to teaching level, 57% of teachers in the sample reported that they taught at the elementary level and 26% indicated that they were secondary teachers. Over one-fourth (29%) had earned bachelors degrees and two-thirds (67%) had earned masters degrees. Finally, 16% reported having 1-5 years of teaching experience, 14% reported having 6-10 years of experience, 17% had 11-15 years, 12% had 16-20 years, 22% reported having 21-25 years, and 18% indicated that they had more than 25 years of teaching experience.

The sample of preservice teachers consisted of 67 students. The only demographic information available for this group consisted of the gender of each student, as the participants would not have been able to respond to the other previously listed demographic items such as educational level, years of experience, and so on. Forty-three percent of the preservice sample was female; 57% was male. The demographic characteristics for both groups are summarized in Table 2.

It is important to note that, although the sample sizes for both groups were not large, the demographic characteristics of each as reported here very closely resemble those of the entire population of teachers not only in the three county region, but also in the entire state of Ohio, as reported by the Ohio Department of Education. Therefore, it could be assumed that the two groups of teachers did in fact constitute representative groups.

Table 3
t-Test Results for Comparisons of Scores for Preservice[a] and Inservice[b] Teachers

Standard	Group	Mean[c]	*t*-statistic	*p*-value
Standard 1				
Choosing Appropriate Assessment Methods	Preservice	3.25	3.79*	<.001
	Inservice	3.74		
Standard 2				
Developing Appropriate Assessment Methods	Preservice	2.78	3.28*	.001
	Inservice	3.18		
Standard 3				
Administering, Scoring, and Interpreting the	Preservice	3.24	5.23*	<.001
Results of Assessments	Inservice	3.95		
Standard 4				
Using Assessment Results to Make Decisions	Preservice	2.67	4.36*	<.001
	Inservice	3.36		
Standard 5				
Developing Valid Grading Procedures	Preservice	2.06	–.03	.975
	Inservice	2.06		
Standard 6				
Communicating Assessment Results	Preservice	2.27	1.69	.093
	Inservice	2.57		
Standard 7				
Recognizing Unethical or Illegal Practices	Preservice	2.69	2.77*	.007
	Inservice	3.10		
Total Score	Preservice	18.96	4.85*	<.001
	Inservice	21.96		

[a] $n = 67$
[b] $n = 197$
[c] The mean score for each *Standard* ranges from a possible low score of 0 to a high score of 5 (indicating the average number of items per *Standard* answered correctly).
* $p < .01$.

Descriptive results for preservice teachers

Data resulting from the preservice teacher group (N = 67) demonstrated a reasonably good level of internal consistency reliability, α = .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 on each of the seven standards are presented in Table 3.

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*.

Descriptive results for inservice teachers

Data resulting from the inservice teacher group (N = 197) demonstrated a mediocre level of internal consistency reliability, α = .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 on each of the seven standards are also presented in Table 3.

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*.

Comparative results for the two groups of teachers

Standard and total scores for the two groups of teachers were compared by conducting independent-samples t-tests (α = .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, meaning 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 shown in Table 3.

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 (α = .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 (α = .74 for both the original study and the study at hand).

Comparisons between preservice and inservice teachers of the seven competency area scores revealed signifi-

cant 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*. It is reasonable to expect that practical experience with student assessment in classroom settings would result in teachers possessing greater knowledge of and superior abilities to apply various assessment terms and concepts, as compared to their preservice counterparts. Participants' performances in this study on five of the seven standards, as well as on the total *CALI* score, support this assertion. However, it is a bit alarming that the inservice teachers did not demonstrate this expected higher level of understanding and application skills on two of the Standards, namely Standard 5—*Developing Valid Grading Procedures* and Standard 6—*Communicating Assessment Results*. These are two very critical Standards—so much so that Brookhart, in two of her papers, chose to focus on improving the instruction provided to preservice teachers on only these two competencies (1998, 1999b). She believes that instruction provided in these areas are typically "simplified psychometric content" as opposed to the *application* of those concepts to what teachers are actually called upon to do in their classrooms. Therefore, teachers are not taught how to apply the theories and principles behind valid grading procedures and communication of results to the classroom setting.

Another possible reason—somewhat related to the first—for this lack of difference between the two groups may be due in part to the fact that both Standards address knowledge and skills that even the most experienced teachers struggle with. For example, a portion of Standard 5 states:

> *Teachers will understand and be able to articulate why the grades they assign are rational, justified, and fair, acknowledging that such grades reflect their preferences and judgments. Teachers will be able to recognize and to avoid faulty grading procedures such as using grades as punishment. They will be able to evaluate and to modify their grading procedures in order to improve the validity of the interpretations made from them about students' attainments.*

Brookhart (1993) studied teachers' grading practices and discovered that teachers apply grading scales differently for students depending on their ability levels. She also found out that many teachers continue to award missing work a grade of zero, indicating punitive consequences, even if it meant that a student would fail a course. Furthermore, she concluded that teachers' grading is often a miscellany of attitude, effort, and achievement (1993), and these factors may not always be applied equally across the board to grades assigned to students. With respect to this lack of difference in performance between preservice and inservice teachers in this study, it could be the case that competencies related to grading systems and communicating assessment results are not acquired through practice and experience in the manner that some other competencies such as selecting appropriate assessment methods or developing appropriate assessment methods.

It is important to recognize that the low reliability coefficients—especially that for the group of inservice teachers—serves as a substantial limitation to the results of this study. An apparent lack of reliability in the data resulting from the administration of this particular instrument limits the extent to which the results of this study may be generalized to other groups of both preservice and inservice teachers. At a minimum, it is recommended that the *CALI* be substantially revised—if not completely rewritten—prior to being used in future research studies as a means of measuring teachers' assessment literacy.

Although these low reliabilities are somewhat problematic in terms of generalizing the results of this study, it is also imperative to recognize that the *CALI* was merely a *slightly* modified version of a previously utilized instrument. However, these slight modifications did not result in meaningful—and, in some cases, *any*—differences between the psychometric qualities of the original and revised versions of the instrument. With respect to measuring preservice teachers' assessment literacy, the original instrument and its revised version resulted in identical values for internal consistency reliability. The reliability resulting from the inservice teachers' data in this study was somewhat lower than that for the comparable group of teachers in the original study.

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). It is likely that one course in assessment and measurement may truly be insufficient to cover everything that secondary teachers need to know. The traditional focus of these teacher prep assessment courses has historically been on large-scale standardized testing (Schafer, 1993), although this trend is changing. This changing trend is evidenced by Popham's (2000) call to stop the "erroneous and educationally harmful appraisal of instructional quality via standardized tests…" (p. 15). Further evidence can be gleaned through a brief examination of older and newer classroom assessment textbooks. Older textbooks—for example, Ebel and Frisbie, 1991, and Hopkins, 1998—tend to contain more chapters on standardized testing (3 of 18 chapters, and 3 of 15 chapters, respectively) and fewer on classroom assessment techniques, and in particular, methods of alternative assessment (1 of 18 chapters, and none of 18 chapters, respectively). Newer textbooks on classroom assessment demonstrate a reversal of this tendency. For example, McMillan's (2001a) textbook contains 1 of 13 chapters on standardized testing and 4 chapters on alternative assessment. Similarly, Mertler's (2003) text includes 1 of 13 chapters on standardized testing and 3 chapters on alternative assessment techniques. However, it is also important that the current administration's emphasis on standardized testing, as out-

lined in the *No Child Left Behind Act of 2001* (U.S. Department of Education, n.d.), not be overlooked. Teachers must be proficient in all of these areas of assessment.

The fact that courses in classroom assessment are not well matched with what teachers need to know for classroom practice 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 parents, policymakers, practitioners, teachers, administrators, and counselors. 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.

Recommendations

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; 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.

Based on the findings of this study, as well as on the questions posed above, several recommendations for practice and research are offered here. It is the belief of this author that assessment training at *both* the preservice and inservice levels is crucial. Additionally, this belief is not meant to take away from the valuable knowledge and skills gained through practical classroom experience. Therefore, an initial recommendation is that, although the appropriateness of preservice training in classroom assessment was questioned above, it is certainly not being advocated that the profession abandon this training. On the contrary, preservice training of teachers in the concepts and techniques of classroom assessment is critical. This should be enhanced through thoughtful examination and research into the knowledge and skills that these teachers will need to possess once they assume the responsibilities for their own classrooms and students.

Second, even though assessment training for preservice teachers is important, ongoing training on various topics related to classroom assessment should be an essential component of any district's program of professional development for its teachers. Administrators at both the district and individual building levels need to stress to their teachers the

importance of sound assessment practice and the professional benefits of being assessment literate. Furthermore, they must provide sufficient opportunities for those teachers to improve their understanding and application of assessment techniques.

Third, future research should investigate various reasons behind the apparent discrepancy between the assessment literacy of preservice teachers and that of inservice teachers. The inservice teachers in this study appeared to be significantly more literate than their preservice counterparts with respect to (1) administering, scoring, and interpreting the results of assessments, and (2) using assessment results to make decisions. Additionally, the inservice teachers scored highest in the skill area of administering, scoring, and interpreting the results of assessments; whereas, the preservice teachers scored highest on their abilities to choose appropriate assessment methods. Examination of these differences and the relative impact of preservice training versus "on-the-job" learning certainly seems warranted.

Finally, the measurement community must take on the responsibility of improving assessment literacy among *all* educational stakeholders. These stakeholders include—but are not limited to—administrators, teachers, parents, policymakers, journalists, and the general public. The ability to assess student performance—and to do so in appropriate, valid, and reliable ways—is arguably one of the most important aspects of the job of teaching.

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. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, *30*, 123-142.

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, DC: 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/

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (8th ed.). Boston: Allyn & Bacon.

McMillan, J. H. (2001a). *Classroom assessment: Principles and practice for effective instruction* (2nd ed.). Boston: Allyn & Bacon.

McMillan, J. H. (2001b). 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. (2003). *Classroom assessment: A practical guide for educators*. Los Angeles: Pyrczak.

North Central Regional Educational Laboratory. (n.d.). *Indicator: Assessment*. Retrieved July 24, 2003, from http://www.ncrel.org/engauge/framewk/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. (2000). The mismeasurement of educational quality. *The School Administrator*, *11*(57), 12-15.

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.

U.S. Department of Education. (n.d.). *The facts about measuring progress—No Child Left Behind*. Retrieved June 12, 2003, from http://www.nclb.gov/start/facts/testing.html

Reviewers for 2005

Savilla Banister
Bowling Green State University

Tom Cody
Western Illinois University

Sharon Damore
DePaul University

Cathleen Doheny
University of West Georgia

Ros Duplechain
University of West Georgia

Mark Earley
Bowling Green State University

Eva Floyd
University of Kentucky

Cynthia P. Fraga-Cañadas
The Ohio State University

Myrna Gantner
University of West Georgia

Cher Hendricks
University of West Georgia

Marie Holbein
University of West Georgia

Sema A. Kalaian
Eastern Michigan University

Barbara Kawulich
University of West Georgia

Michael Kerstetter
The Ohio State University

Jenny Kilgore
Miami University

Charles E. Kline
Purdue University

Dorothy Kupsky
The Ohio State University

Terri Lane
University of Dayton

Bambi Bethel Leitshuh
University of Missouri-St. Louis

Robert H. Lombard
Western Illinois University

Eleni Makris
Northeastern Illinois University

Gail Marshall
University of West Georgia

Julia M. Matuga
Bowling Green State University

Sharon McNeely
Northeastern Illinois University

Harry Morgan
University of West Georgia

Glenda Moss
Indian University-Purdue University Fort Wayne

Timothy James Murnen
Bowling Green State University

Noriko Nagata
The Ohio State University

Sungworn Ngudgratoke
Michigan State University

Betul Ozkan
University of West Georgia

Lynn M. Papenfus
Kent State University

Jonnie Jill Phipps
University of Akron

Jean Pierce
Northern Illinois University

Hema Ramanathan
Butler University

Ronald S. Reigner
University of West Georgia

Laura Renzi
The Ohio State University

Elaine Roberts
University of West Georgia

Kenneth Royal
University of Kentucky

Mary Beth Slone
University of West Georgia

Ruslan Slutsky
University of Toledo

Janet Strickland
University of West Georgia

Kathleen Sullivan-Brown
University of Missouri

Sharon Valente
Ashland University

John vonEschenbach
University of West Georgia

Kathryn Wiggins
DePaul University

E. Jane Williams
The Ohio State University

James P. Wilson
University of Dayton