

2000

## Teacher-Centered Fallacies of Classroom Assessment Validity and Reliability

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. (2000) "Teacher-Centered Fallacies of Classroom Assessment Validity and Reliability," *Mid-Western Educational Researcher*. Vol. 13: Iss. 4, Article 8.

Available at: <https://scholarworks.bgsu.edu/mwer/vol13/iss4/8>

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.

---

# *Teacher-Centered Fallacies of Classroom Assessment Validity and Reliability*

Craig A. Mertler  
Bowling Green State University

## *Abstract*

*The general purpose of this descriptive study was to examine the current assessment practices of teachers in the state of Ohio. Specifically, the aim of this study was to gain an understanding of the methods used to insure the validity and reliability of their classroom assessments. It builds on previous research by incorporating verbal explanations and categorizations of techniques used to insure classroom assessment validity and reliability. Similar to previous research, it was determined that teachers do not spend much time conducting statistical analyses of their assessment data. Teachers seemed to have a better understanding of assessment reliability than validity; although, many of the steps provided by teachers to determine classroom assessment validity and reliability would be considered poor and inadequate, at best. Recommendations include providing additional inservice training to teachers, as well as tailoring measurement courses to fit the needs of future classroom teachers.*

## Background

A sizable amount of classroom time is devoted to the assessment of student learning. Since teachers must give even more time to the preparation and scoring of tests and other assessments, a substantial proportion of a teacher's day is devoted to issues surrounding student assessment. One could argue, then, that careful consideration of testing within formal teacher preparation programs is certainly warranted. If educators, particularly those in teacher preparation programs, are to help teachers use their student testing time efficiently and to be effective at it, more must be learned about how teachers perceive and use classroom tests and other forms of assessment (Gullickson, 1984).

Several research studies examining the overall assessment practices of classroom teachers have been conducted; however, little research on the topic of practices with respect to insuring classroom test validity and reliability exist in the literature. Much of the research has focused on the use of various types of items and differences that exist across school levels (i.e., elementary, middle, and high schools) and school locations (i.e., urban, suburban, and rural). For example, Marso (1985; 1987) found several differences between elementary and secondary teachers. Secondary teachers tended to use more self-constructed tests rather than published tests; whereas, the opposite was true for elementary teachers, especially those in grades K-4. Similarly, others have found that the higher the grade level, the greater the tendency for teachers to use their own assessments (Stiggins and Bridgeford, 1985). Secondary teachers reported relatively more use of essay and problem-type items and less frequent use of completion and multiple-choice items than did elementary teachers (Marso and Pigge, 1987). Marso (1985) also found that teachers perceived matching, multiple-choice, and completion type items as being most useful.

Establishing the validity of classroom assessments has undergone a recent shift in focus. In the past, measurement experts wrote about three types of validity: content, construct, and criterion. The most important of these for the classroom teacher was arguably content validity—the extent to which the content of a test or other assessment activity corresponds to the performance to be observed (Oosterhof, 1999). Prior planning was a key to establishing a test's content validity, and this planning typically consisted of the careful development of performance objectives. These performance objectives not only guided instruction, but also served as the catalyst for the development of actual items or activities that made up the assessment (Oosterhof, 1999). A valid assessment is one that provides students with the opportunity to show what they have learned following instruction (Airasian, 2000). Therefore, when developing a valid assessment, teachers should focus attention on instructional objectives as well as the actual instruction that took place, and should do so *during* the development of the assessment (Airasian, 2000; Oosterhof, 1999).

More recently, however, validity is seen as a dynamic concept, referred to as “construct validity” (Gredler, 1999), that incorporates all three—previously separate—types of validity. Gredler (1999) advises that teachers should ask themselves a series of questions regarding their classroom assessments as a means of determining validity:

- *Does the item or task match the instructional method used?*
- *Does the item or task relate directly to the class objectives?*
- *Can all students who understand the concepts demonstrate their knowledge with the particular assessment?*

Even with this revised view of validity, teachers are still advised to establish classroom assessment validity by means of professional judgment (Airasian, 2000; Gredler, 1999;

McMillan, 1999) and by comparing the items or tasks to instructional methods and objectives.

Establishing the reliability of classroom assessments is a more structured—and decidedly, more objective—process. There are six statistical methods which can be used, depending on the specific situation and what type of consistency information is desired (Gredler, 1999). Reliability can be established for assessments that yield a range of scores through one of the following methods: test-retest, equivalent (or alternate or parallel) forms, split-half, Kuder-Richardson (KR-20), or coefficient alpha (Gredler, 1999; Oosterhof, 1999). Since most classroom assessment activities are administered only once, and typically consist of right or wrong answers, the split-half and KR-20 methods are most appropriate. If performance or portfolio assessments are used, a percentage agreement between raters can be calculated (Gredler, 1999; Oosterhof, 1999).

For years, measurement experts have told us how teachers *should* establish classroom assessment validity and reliability. However, very little empirical information exists on how teachers *actually* determine the extent to which their assessments are valid and reliable. However, some research on teachers' use of statistical analyses of test data does exist. Several studies have documented the infrequent use of statistical analyses of test data (Gullickson, 1986; Marso and Pigge, 1987; Marso and Pigge, 1988). This may be due to the fact that teachers are not convinced of the value of using statistical procedures to improve the quality of their tests or that they simply do not have a good grasp of statistical concepts and this discomfort may lead to a devaluing of their use.

This study was part of a larger research endeavor which had as its main purpose the examination of the current assessment practices of K-12 teachers in the state of Ohio. The researcher sought to explore how practicing teachers assess student performance with their students in their own classroom settings. Specifically, the goal of this research study was to gain an understanding of the processes and techniques used by classroom teachers to insure that their assessments are both valid and reliable, and to determine the extent to which they engage in these processes.

## Methodology

The researcher made use of resources available through the Ohio Department of Education in order to obtain a stratified random sample of K-12 teachers throughout the state of Ohio. The sample was stratified so that various subgroups in the population of K-12 teachers in the state were represented in the sample in the same proportion that they exist in the population. These subgroups of teachers included the following four categories: (1) female elementary, (2) female secondary, (3) male elementary, and (4) male secondary. A random sample of 3,000 teachers was obtained.

An original survey instrument, the *Ohio Teacher Assessment Practices Survey*, was developed by the researcher

for purposes of collecting the data. The literature was relied upon heavily in order to guide the development of the specific items appearing in the survey instrument. The instrument consisted of 47 items and included both scaled (forced-choice) and open-ended items. For purposes of the study at hand, teachers were asked to respond to items concerning the validity and reliability of their classroom assessments, specifically requesting information on the steps that they follow and the extent to which they do so.

In mid-January, each teacher received a packet containing a full-page cover letter, copy of the survey, and a self-addressed, postage-paid return envelope. They were instructed to return the survey within four weeks from the date appearing on the cover letter. In mid-February, a follow-up reminder postcard was sent to those teachers who had not yet returned completed surveys. The final sample upon which the analyses were conducted consisted of 625 completed surveys. Analyses were conducted using SPSS (v. 6.1) and NUD\*IST (v. 4).

It should be noted that the 21% response rate may initially seem problematic, especially with respect to the generalizability of results. However, two important points can justify their generalizability. First, Gay and Airasian (2000) state that once a population surpasses approximately 5,000 members, its "size is almost irrelevant and a sample size of 400 will be adequate" in order for the researcher to be confident in the generalizability of the results (p. 135). Based on this fact, this study's stratified random sample of  $n = 625$  teachers is representative of the more than 100,000 teachers in the state of Ohio. Second, in order to insure representativeness, the researcher compared general respondent characteristics in the sample to those in the entire target population, utilizing data obtained from the Ohio Department of Education web site (<http://www.ode.state.oh.us/>). Since the obtained sample was based on proportional representation within the four subgroups previously listed above, comparisons were made to the analogous proportions within the target population. The result of this informal comparison is presented in Table 1. The proportion of representation within the four

Table 1  
*Comparison of sample and population characteristics by frequencies (and percentages)*

Demographic Characteristic	Sample (n = 625)	Population (N = 101,092)
School Level by Gender		
Elementary – Female <sup>a</sup>	114 (70%)	56,160 (82%)
Elementary – Male <sup>b</sup>	50 (30%)	12,703 (18%)
Secondary – Female <sup>c</sup>	158 (45)	16,868 (52%)
Secondary – Male <sup>d</sup>	191 (55%)	15,361 (48%)
Years of Teaching Experience <sup>e</sup>		
1-5 Years	84 (14%)	17,879 (18%)
6-10 Years	103 (17%)	15,184 (15%)
11+ Years	434 (70%)	63,487 (63%)

<sup>a</sup>Calculated as the percentage of elementary teachers who are female

<sup>b</sup>Calculated as the percentage of elementary teachers who are male

<sup>c</sup>Calculated as the percentage of secondary teachers who are female

<sup>d</sup>Calculated as the percentage of secondary teachers who are male

<sup>e</sup>Calculated as the percentage of the total sample or population

subgroups is fairly similar, with the larger discrepancy occurring between males and females at the elementary level. As is also shown in the table, there exists a great deal of similarity between the sample and population with respect to years of teaching experience. Based on this combined information, it was concluded that the resultant sample findings could indeed be generalized to the population of Ohio teachers.

### Results

The sample consisted of 53% females and 47% males. The majority (42%) of teachers were from suburban settings, followed closely by rural (32%) and urban (25%). Nearly half (47%) were teaching at the senior high level; just over one-fourth (26%) were teaching at the elementary level, followed closely by those teaching at the junior high/middle school level (25%). Twenty percent of the teachers had 26-30 years of teaching experience, followed by 21-25 years (19%), 6-10 years (17%), 1-5 years (13%), 16-20 years (13%), 11-15 years (11%), and 31-35 years (6%). Two teachers in the sample had 36 years or more of teaching experience.

#### Validity of Classroom Assessments

Teachers were asked to list specific steps they followed to insure that their assessments were valid and to indicate how often they followed these steps. Specifically, they were asked to respond to the following open-ended question:

*“What specific steps should teachers follow to make sure their written tests or other assessments*

*are valid (that is, actually measure what students have learned)?”*

Following their responses to this item, teachers were then asked to respond to the following:

*How often:*

a. *are you able to closely follow these steps?*

1 2 3 4 5

b. *do you believe teachers closely follow these steps?*

1 2 3 4 5

where 1 = never, 2 = not very often, 3 = about half of the time, 4 = most of the time, and 5 = always. One-fourth (25%) of the teachers responded that they followed specific steps to insure validity about half of the time or less; the median response was “most of the time.” Two-thirds (66%) of the teachers believed that teachers, in general, followed those steps about half of the time or less; the median response was “about half the time.”

With respect to the specific steps that teachers follow to insure validity, a wide variety of responses was provided. Six hundred and eleven responses were examined and categorized based on common approaches. The resulting hierarchical coding system is shown in Figure 1.

The teachers’ responses were coded into six major categories, with the vast majority falling into roughly two of those categories. The major categories, with the numbers and percentages of response appearing in parentheses, were as follows:

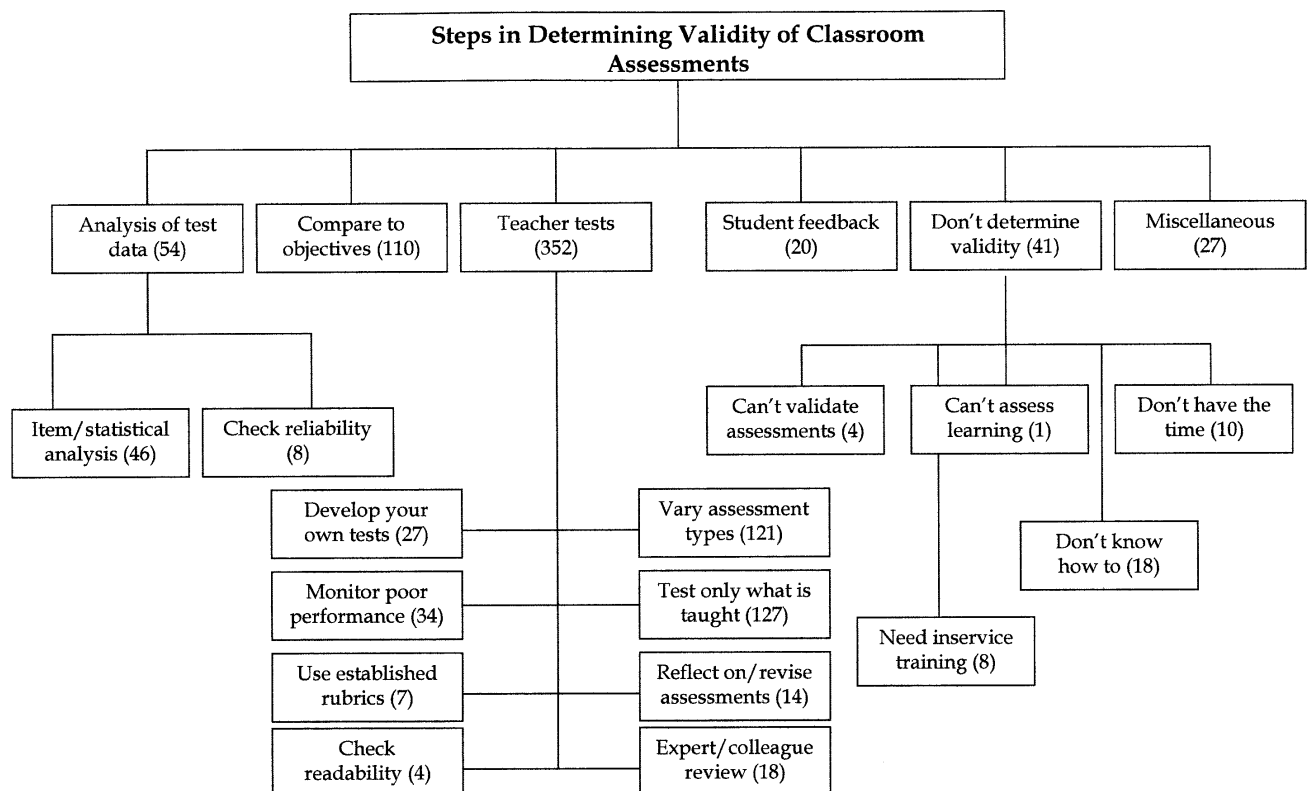


Figure 1. Coding scheme for teachers’ approaches to determining classroom assessment validity.

- teacher-developed tests (352 or 58%);
- compare to objectives or curriculum (110 or 18%);
- analysis of test data (54 or 9%);
- don't determine validity (41 or 7%);
- ask for student feedback (20 or 3%); and
- miscellaneous (27 or 4%);

Several of these major categories included anywhere from a couple to several sub-categories. The sub-categories, along with the frequencies of response, are provided in Figure 1.

As evident in Figure 1, more than half of the responses dealt with teacher-made tests. The vast majority of teachers stated that they insure assessment validity by following conventional rules of sound test development, varying the types of items and assessments (thus providing students different means of showing what they know), and by simply testing what is taught. Several comments exemplifying these points follow:

*Vary the type of questions in terms of difficulty and questioning types.*

*Use essay questions, not multiple guess or True or False...I did not know what students understood giving multiple choice and True or False tests. Essays are more time consuming to grade, but well worth the effort.*

*Make sure all tests are varied enough in questioning to accommodate all learning styles, not just one or two styles of questions.*

*Written tests should be based entirely on what was taught.*

Many teachers believe that simply developing your own assessments, as opposed to using published materials, will insure assessment validity. Other teachers tend to monitor their students' performance on their self-developed assessments; if students perform poorly, they make adjustments accordingly. One teacher stated

*I take the tests as I go. If there are questions that most students bomb, I'll eliminate it, but if I feel they were well prepared for it, I'll keep the question.*

Several teachers believe that simply reflecting on the success of an assessment instrument, evaluating how well students performed, and then revising the instrument would insure validity. Along these lines, teachers suggested asking questions of the students in order to gather feedback concerning the assessment. For example,

*I...have them evaluate the test according to what I taught or thought I taught.*

Finally, with respect to teacher-made assessments, a small sampling stated that they have "experts" or other teachers review their tests and other assessments as a means of checking the validity:

*My colleagues and I pass tests around to each other to see if everyone is on the same level.*

Many teachers insure validity by comparing their assessments to instructional objectives or the district/statewide curriculum.

*Compare assessment to objectives in order to evaluate individual questions.*

*Ask questions based on the material to be learned/course of study/curriculum...try to see what they know as well as what they don't know.*

Many teachers rely on the results of statistical analyses of test data or other information resulting from assessments. Several teachers stated that they simply "checked reliability" as a means of insuring validity, without providing any details of how they did so. Others specifically stated that they conducted item analyses of student data, although their approaches to doing so may have been a little vague:

*Use statistics to validate the reliability.*

A small proportion of teachers stated that they didn't attempt to validate their assessments for a variety of reasons including the fact that validation cannot be done, student learning cannot truly be assessed, and there just is not enough time to do so. However, the majority of teachers who responded in this category confessed that they simply did not know how to validate their assessments and that inservice training was desperately needed.

*Get professionals to inservice with applications for practical use. Experiment with these methods. Choose the methods which best fit the specific needs.*

*Teachers need concrete examples and explicit instruction on how to create valid assessment items for written tests.*

Miscellaneous comments covered a wide range and encompassed several areas not covered by the broad categories. These included comments related to comparisons to proficiency test scores, the issue of cheating, and taking the test yourself to see if it appears valid.

It is clear that, although many of these comments provide sound advice for teachers to follow, these "steps" simply are not appropriate—or are incomplete and lack thoroughness—for determining the validity of classroom assessments. By following good test development guidelines, teachers will certainly be more likely to achieve tests that are valid, but simply following those rules will not insure validity. Several teachers seemed to have the concept of reliability confused with that of validity when they identified item analyses as a means of validation.

For many classroom assessments, content validity would be the most important type of validity to establish. Unfortunately, less than 20% of the teachers' comments focused on specific comparisons of assessment items and activities to instructional objectives, although another 21% of the com-

ments identified the matching of assessments to what was actually taught. The idea of simply using self-developed tests and varying the types of assessments alone is not enough to insure validity. Careful planning of this type certainly helps with assessment validity, but it must be accompanied by the establishment of congruency with objectives.

It should be noted that several teachers provided miscellaneous comments that definitely could *not* be considered means of establishing validity and appeared to be somewhat troublesome. These included:

*Although my techniques are not written down any longer, I use a mental format which I change as needed. Experience is a wonderful resource.*

*Over the years, you'll find out what works for you.*

*It takes me over an hour to even write a new test. To be honest, other than using my experience, I don't have much time to worry about how valid my test is.*

*I don't know. Most of the time I am so busy I don't have time to check validity. I guess I leave this job up to someone else.*

*No clue! I have no training is doing this, and never really thought about it until reading this question.*

*Teachers don't have time for this type of analysis! Why don't you teach in a public school for a year and find out what it is really like.*

### Reliability of Classroom Assessments

Teachers were also asked to list specific steps they followed to insure that their assessments were reliable and to indicate how often they followed these steps. Specifically,

they were asked to respond to the following open-ended question:

*“What specific steps should teachers follow to make sure their written tests or other assessments are **reliable** (that is, **consistently** measure what students have learned)?”*

Following their responses to this item, teachers were again asked to respond to the following:

*How often:*

a. *are **you** able to closely follow these steps?*

1    2    3    4    5

b. *do you believe **teachers** closely follow these steps?*

1    2    3    4    5

where 1 = never, 2 = not very often, 3 = about half of the time, 4 = most of the time, and 5 = always. Nearly one-third (30%) of the teachers responded that they followed specific steps to insure reliability about half of the time or less; the median response was “most of the time.” Two-thirds (66%) of the all teachers believed that teachers followed those steps about half of the time or less; the median response was “about half the time.”

When asked to provide the specific steps that they follow to insure reliability, the teachers again provided a wide variety of responses. Four hundred and thirty-one responses were examined and categorized based on common approaches. The resulting coding system is shown in Figure 2.

The teachers' responses were coded into five major categories, with the vast majority falling into one of those categories. The major categories, with the numbers and percentages of response appearing in parentheses, were as follows:

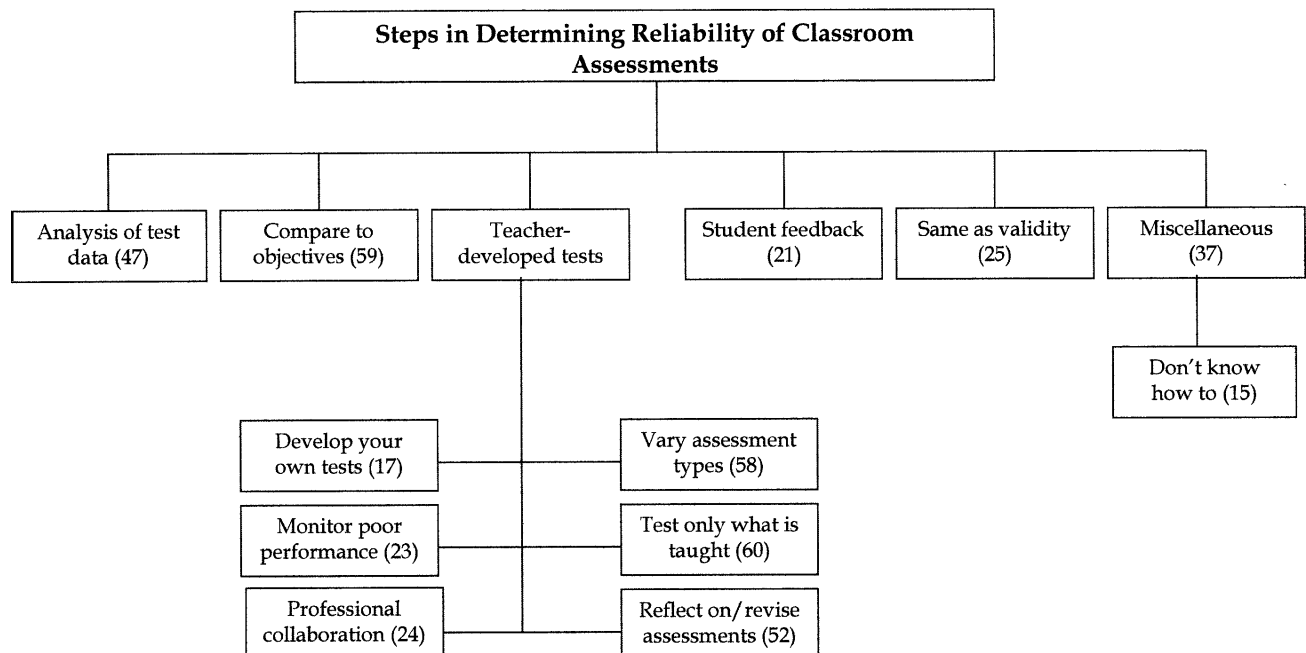


Figure 2. Coding scheme for teachers' approaches to determining classroom assessment reliability.

- teacher-developed tests (234 or 54%);
- compare to objectives or curriculum (59 or 14%);
- analysis of test data (47 or 11%);
- same process as validity (25 or 6%);
- ask for student feedback (21 or 5%); and
- miscellaneous (37 or 9%);

Several of these major categories included anywhere from a couple to several sub-categories. The sub-categories, along with the frequencies of response, are provided in Figure 2.

As is evident from Figure 2, many teachers believe that insuring assessment validity and reliability are very similar procedures. Many of the same coding categories emerged as a result of examination of the responses to question addressing classroom assessment reliability. Again, the majority of teachers stated that they insure assessment reliability by following conventional rules of sound test development, varying the types of items and assessments, and by simply testing what is taught.

Similar to the responses regarding validity, teachers tend to monitor their students' performance on the self-developed assessments and revise them accordingly, as well as gathering oral feedback from students themselves about the assessment instruments or activities.

Again, many teachers identified professional collaboration as a means of insuring reliability, as well as comparing assessments to instructional objectives. Unfortunately, few teachers (11%) rely on the results of statistical analyses of test data or other data resulting from assessments to insure reliability. However, several teachers explicitly stated that they utilized "test-retest" or "equivalent forms" methods of determining the extent to which their assessments are reliable.

A small proportion of teachers again stated that they did not know how to demonstrate the reliability of their assessments and that inservice training was necessary.

Miscellaneous comments included those related to comparisons to proficiency test scores, a teacher's knowledge of the content, performing readability tests on assessment instruments, and establishing a consistent grading system.

It seems that many teachers have a slightly better grasp of the concept of validity than that of reliability, especially in terms of establishing those characteristics for their classroom assessments. However, the overriding majority of comments provided would not be considered acceptable methods of determining either classroom assessment validity or reliability.

It should be noted that several teachers provided miscellaneous comments that should again "raise a red flag" concerning their knowledge and ability to appropriately assess reliability. These included:

*...techniques such as test-retest are possible, but they aren't practical in day to day classroom.*

*Check grades...compare scores with what was taught. Use common sense.*

*I would determine the percentage of students who demonstrate the ability you're looking for. Determine a ranking (90% answer correctly, then it is reliable).*

*No specific steps. There are too many other things required of teachers.*

*I really don't understand the difference between validity and reliability...sorry! Is it just me?*

*...with all the other tasks at hand, worrying about the reliability of my tests is way down at the bottom of my priority list. I use my experience to determine reliability...*

*What's the difference between reliable and valid -- really?*

## Conclusions

This study was part of a larger research endeavor which had as its main purpose the examination of the current assessment practices of K-12 teachers in the state of Ohio. Specifically, the goal of this research study was to gain an understanding of the processes and techniques used by classroom teachers to insure that their assessments are both valid and reliable, and to determine the extent to which they engage in these processes.

This study was successful in that it resulted in a somewhat thorough description of these teachers' assessment practices with respect to issues of validity and reliability of their classroom assessments. It builds on previous classroom assessment practices research by incorporating information about validity and reliability analyses, which is quite scarce. Similar to previous research, it was determined that teachers do not spend much time conducting statistical analyses of their assessment data.

Previous research has shown that many teachers do not believe that they are well prepared to assess student performance. Mertler (1999, 1998) asked teachers to indicate their level of preparation—in terms of assessing student learning—that resulted from their undergraduate teacher education program. Teachers were asked to respond on a five-point scale, where 1 = not at all prepared, 2 = not very prepared, 3 = slightly prepared, 4 = somewhat prepared, and 5 = well prepared. The median response from the more than 600 teachers was "slightly prepared," with only 13% indicating that they felt "well prepared." Similarly, Quilter and Chester (1998) reported that many teachers in their study admitted that their training in testing and measurement is somewhat deficient.

The results of this study, coupled with previous research, perhaps imply that some attention needs to be re-focused on

undergraduate teacher preparation measurement courses, especially in the areas of validity and reliability. Although these teachers *claim* they do a good job of following steps to insure sound assessments, they do not possess a solid foundation of what those steps should be. In other words, they frequently evaluate validity and reliability, but do so in the wrong ways. Therefore, they are really not evaluating those critical characteristics of classroom assessments. To further complicate this problem, the participating teachers believed that they use these techniques—albeit, the wrong techniques—more frequently than most other teachers. Only when measurement courses provide solid foundational understanding of these concepts will we have adequately prepared our teachers to assess their students' performance.

However, it may be more appropriate to focus teaching and training efforts on inservice—rather than preservice—teachers. McMillan (1999) stresses the importance of training and other opportunities that allow teachers to “brush up on their assessment skills.” He continues by stating that teachers are “simply expected to be able to administer most any kind of assessment without adequate training...” Others (Quilter and Chester, 1998) have also cited implications for inservice training in educational assessment—specifically to help teachers see the *value* in the appropriate use of their various approaches to assessment, instead of simply showing them how to do the assessment. It may be the case that teachers in general need to and should have some teaching and assessment experience—beyond the training received during undergraduate coursework and student teaching—prior to being able to completely understand the concepts of validity and reliability, be able to consider those concepts during the development of their classroom assessments, and be able to appropriately assess these characteristics. Professional development in the form of inservice training is definitely something that numerous teachers in this study identified as being necessary, useful, and needed.

#### Acknowledgement

This research was made possible by a grant from the Martha Holden Jennings Foundation, Cleveland, OH (Grant No. A-107-97). The author wishes to express his appreciation to the Foundation for its support.

#### References

- Airasian, P. W. (2000). *Assessment in the classroom: A concise approach (2<sup>nd</sup> ed.)*. Boston: McGraw-Hill.
- Gay, L. R. and Airasian, P. (2000). *Educational research: Competencies for analysis and application (6<sup>th</sup> ed.)*. Upper Saddle River, NJ: Merrill.
- Gredler, M. E. (1999). *Classroom assessment and learning*. New York: Longman.
- Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests. *Journal of Educational Research*, 77(4), 244-248.
- Gullickson, A. R. (1986). Teacher education and teacher-perceived needs in educational measurement and evaluation. *Journal of Educational Measurement*, 23(4), 347-354.
- McMillan, J. H. (1999, March). *Establishing high quality classroom assessments*. Metropolitan Educational Research Consortium, Virginia Commonwealth University. (ERIC Document Reproduction Service No. ED 429 146)
- Marso, R. N. (1985, October). *Testing practices and test item preferences of classroom teachers*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 268 145)
- Marso, R. N. and Pigge, F. L. (1987, October). *Teacher-made tests: Testing practices, cognitive demands, and item construction*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA. (ERIC Document Reproduction Service No. ED 298 174)
- Marso, R. N. and Pigge, F. L. (1988, October). *An analysis of teacher-made tests and testing: Classroom resources, guidelines, and practices*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 291 781)
- 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. (1998, October). *Classroom assessment practices of Ohio teachers*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. 428 085)
- Ohio Department of Education. (2000). *Ohio Department of Education Home Page*. [WWW document, <http://www.ode.state.oh.us/>]
- Oosterhof, A. (1999). *Developing and using classroom assessments (2<sup>nd</sup> ed.)*. Upper Saddle River, NJ: Merrill.
- Quilter, S. M. and Chester, C. (1998, October). *Inservice teachers' perceptions of educational assessment*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 425 152)
- Stiggins, R. J. and Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22(4), 271-286.