DOCUMENT RESUME

ED 390 937 TM 024 562

AUTHOR Oosterhof, Albert
TITLE An Extended Observation of Assessment Procedures Used by Selected Public School Teachers.
PUB DATE Apr 95
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Classroom Techniques; Elementary School Teachers; Elementary Secondary Education; *Informal Assessment; Integrated Activities; Observation; Public Schools; *Public School Teachers; Secondary School Teachers; *Student Evaluation; *Test Use; Validity

ABSTRACT

This paper describes findings from a set of extended observations of 15 public school teachers in an effort to gain insight into classroom assessment procedures. Two elementary school, seven middle school, and six high school teachers were observed over 5 days by dividing time into 10-minute intervals. Attention was given to the amounts of time teacher and students spent on five action categories: (1) formal assessment; (2) informal assessment; (3) integrated assessment and instruction; (4) other on-task actions; and (5) off-task actions. The major role of informal assessment and integrated assessment and instruction was in stark contrast to the emphasis provided in current measurement textbooks, suggesting that many textbooks are not responsive to the actual needs of teachers. Validity issues related to the assessments being judged appeared to be linked to the teacher's understanding of the material being taught, but the degree to which the teacher sought criterion-related evidence of validity appeared to be a function of whether the teacher anticipated discrepancies between observed student performance and actual student ability. (Contains three tables and five references.) (SLD)

# An Extended Observation of Assessment Procedures Used by Selected Public School Teachers

Albert Oosterhof
Florida State University

This paper describes findings from a set of extended observations of fifteen public school teachers. A variety of at-school activities were observed including the teacher's preparation, classroom activities, meetings with other teachers and school administrators, and conferences with parents. Most of the time involved in-class activities. Elementary, middle school and high school teachers were involved in these observations.

The purpose of these observations was to gain insight into current classroom assessment practices. Considerable changes in classroom assessment are being encouraged. Relatively new terms such as alternative and authentic assessment are now commonplace in the literature. Curriculum specialists, particularly in language arts, mathematics and science, are now placing considerable emphasis on the role of assessment within instruction. On the other hand, most textbooks concerned with classroom assessment and possibly most college courses that use these books provide limited coverage of these issues. Furthermore, many faculty who teach these courses have limited direct exposure to K-12 classrooms. Within this context, it was anticipated that a series of extended observations would provide useful information.

## Method

A total of fifteen teachers were observed. Two teach in elementary schools, seven in middle schools and six in high schools. The two elementary teachers work within self-contained classrooms, one at Grade 2 and the other at Grade 5. The seven middle school teachers work as teams in two separate schools. Each team works with a common group of students. The students in the respective middle schools are in Grades 6 and 8. The six high school teachers are affiliated with two schools. Two teach English (literature and writing), two teach mathematics, and two teach science. Each of these individuals typically teaches grades 9 through 12.

Teachers participating in this study were diverse in terms of their characteristics and also with respect to the settings in which they work. Particularly at the middle and high school levels, teachers varied from traditional to progressive with respect to both pedagogy and instructional goals. The two groups of middle school teachers differed with respect to how they functioned as a team. One group used their meetings to coordinate all of their classroom activities whereas the other group used the meetings mostly to address problems being experienced with selected students. The schools at which the fifteen teachers are employed serve students from different populations. One middle school and one high school are within a large consolidated rural district. The other high school is in an urban area drawing students from economically low and middle-class neighborhoods. The other middle school is in an inner-city neighborhood. The students attending both elementary schools are mostly from middle-class neighborhoods.

An attempt was made to observe a variety of classroom environments, although the selection techniques were not rigorous. Schools were selected so as to include the characteristics listed above. To make the extended observations feasible, the schools had to be within commuting distance of the observer. No formal attempt was made to select teachers or schools that represented a carefully defined population. Even if more rigorous sampling techniques had been used, the small number of classrooms involved would have posed problems. The extent to which the present observations generalize, therefore, is unknown. However, the extended observations provide what appears to be useful insights related to assessment procedures used by selected public school teachers. Further research will demonstrate whether these observations are more broadly relevant.

A given teacher or small group of teachers was observed for a period of five consecutive days. Observations on a given day encompassed the entire period of time teachers were at the school. For each of the two elementary school teachers, this involved observing the teacher full-time for the duration of the week. For the middle school teachers, one teacher within the team was observed at a time. At one of the middle schools, each teacher taught sections of two subjects, with different combinations of students assigned to each class. At this school, individual teachers were observed in a sequence that permitted observing each teacher work with both subject areas. At the other middle school, five sets of students rotated in sequence among the four teachers (and among classes taught by non-team teachers such as music and physical education). At this school, observations involved one set of students as they rotated among the teachers. At the high schools, generally two classes taught by each of three teachers were observed for the period of a week. This sequence was used at both high schools. At the elementary and middle schools, observations also included planning periods, meetings, and conferences. Observations of the fifteen teachers took place within a three-month interval.

The use of extended observations provided an opportunity to gain insight into a number of teacher and student behaviors, perceptions, and strategies that perhaps go unnoticed when shorter periods are involved. An extended observation provides the observer an opportunity to understand the situation the teacher and students are in. The teacher and students loose the opportunity to contrive the situation. The observer gains the privilege of becoming unobtrusive.

A week prior to initiating an observation, its purpose was discussed with the teacher or group of teachers. Confidentiality was assured. Some of the teachers indicated concern about being observed particularly for an extended period. Ultimately, none of the teachers who were asked to participate refused to be observed, although each clearly had that option. In every case, the teacher became comfortable with the observational process. Several of the teachers on their own initiative stayed well past school hours to discuss their ideas and strategies.

Observations of the fifteen teachers were guided by a set of questions:

A) What techniques do teachers use to assess students? How do these techniques compare to those discussed in textbooks on classroom assessment?

B) Do teachers' assessments appear to be valid? Does observed student performance appear to generalize to performance the teacher did not observe?

C) What conditions and approaches appear most useful for helping teachers learn assessment techniques?

Documentation of observations took two forms. The first involved recording at 10-minute intervals estimates of the amount of time the teacher and students associated with each of the following five categories:

| Category | Description | Examples |
|---|---|---|
| 1. Formal assessment | Assessment activities that used a previously produced instrument or other formally established procedure | A written test or quiz, performance assessment, observational checklist, previously prepared oral questions, or portfolio |
| 2. Informal assessment | Assessment activities that do not use a previously produced instrument | Casual observation, or spontaneous oral questions |
| 3. Integrated assessment and instruction | Activities in which the teacher's instruction and assessment are interactive and inseparable | Dialog where ideas are developed through the teacher's oral questions |
| 4. Other on-task activity | Any on-task activity void of overt assessment by the teacher | Lecturing, watching a video, reading a story aloud without dialog |
| 5. Off-task activity | Any task that appears unrelated to instructional goals | Day dreaming, resting, conversation among students unrelated to class activities, disruptive behavior |

A teacher and students often are involved in separate categories of behavior at the same time. For instance, while students are completing a formal assessment such as a written test, the teacher is typically monitoring student behaviors (informal assessment). Similarly, a teacher may be working interactively with a subset of students (interactive assessment and instruction) while other students work individually with separate material (other on-task activity). Because the teacher and students often are involved in different categories of behavior, amounts of time were recorded separately for the teacher and students. Amounts of time were recorded as an integer number, ranging from 0 to 10 within teach of the five categories. Across categories, the five numbers summed to 10 for each 10-minute interval. For teachers, the recorded number is a judgment at the conclusion of the interval as to how many minutes of the teacher's activities were associated with each of the five categories. For students, the recorded number again is a judgment, but in this case it is aggregated or averaged across students. For example, if during a 10-minute period an average of 20% of the students had completed a written test and were resting while the remaining students continued working on the test, an 8 would be assigned to formal assessment and a 2 would be assigned to off-task activity. Similarly, if during a 10-minute period the teacher was helping students learn a concept by asking questions of individual students, and during this period 60% of the students appear to be engaged in the process (even if not specifically asked a question) but 40% appear to be disengaged and not involved with task related to instructional goals, a 6 would be assigned to interactive assessment and instruction and a 4 would be assigned to off-task activity.

For purposes of this study, it was presumed that teachers would remain on task while in the classroom. In part, this made it easier to be candid with the teacher as to what was being recorded. In reality, teachers did remain on task throughout the observations.

The second form of documentation of observations took the form of narrative records. These records described what the teacher and students did within each 10-minute period. Emphasis was given to behaviors that appeared to address the previously listed questions that guided the observations. Focus was placed on recording a description of the event rather than an interpretation of the event. Interpretations, however, were included when it was anticipated such information would be required in order to later synthesize the descriptions.

## Results and Discussion

Attention is first given to the amounts of time per 10-minute interval a teacher and students devoted to the five categories of activity. Then information drawn from the narrative records is discussed.

*Amounts of Time*

The amounts of time per 10-minute interval that teachers and students devoted to the five categories of activity were aggregated across the two elementary school teachers, and likewise across the middle school teachers and across the high school teachers. It may be convenient to visualize this aggregation as a two-dimensional matrix, where five rows correspond to the five categories of activity defined earlier, and columns correspond to the 10-minute intervals over time. For purposes of summary, the numbers within each row were re-ordered by ranking, so that each row first listed any zeros recorded within a given category, followed by any 1's, then 2's, and so on through any 10's that were recorded within the category. This rank-ordering going horizontally across each row disjoins the relation between columns. That is, prior to the rank-ordering, a given column was associated with a particular 10-minute interval. If, for instance, a 10-minute interval for students involved a 6 associated with interactive assessment and instruction and a 4 associated with off-task activity, the entries in the remaining categories (rows) would be zero. After the rank ordering, the three zero's would move to the left to join any other zeros within their respective categories (rows), whereas the 6 and 4 would be located somewhere to the right within their respective rows of the matrix. After the rank-ordering, a column to the far left of the matrix might contain all zeros, and a column to the far right might contain all 10's.

The listing of ranked-ordered numbers within each row now presents an ordered listing of the estimated amount of time teachers and students spent, within 10-minute intervals of time within each category, such as formal assessment or informal assessment. If a given matrix included 99 columns (in reality they contained more), the 50th column would list the 50th percentile point as to the amount of time a teacher or students were observed to be involved with a particular category of activity. Other columns would correspond to other percentile points. Tables 1 through 3 list the 5th, 25th, 50th, 75th, and 95th percentile points for the five categories of activity that were observed for the elementary, middle school and high school teachers and their students.

Table 1: Distribution of Actions Occurring in the Elementary School Classrooms

| *Actions of the Teacher* | $P_{05}$ | $P_{25}$ | $P_{50}$ | $P_{75}$ | $P_{95}$ |
|---|---|---|---|---|---|
| 1. Formal Assessment | 0 | 0 | 0 | 0 | 0 |
| 2. Informal Assessment | 0 | 0 | 2 | 7 | 8 |
| 3. Integrated Assessment and Instruction | 0 | 0 | 6 | 7 | 8 |
| 4. Other Actions | 1 | 1 | 3 | 5 | 10 |

| *Actions of the Students* | $P_{05}$ | $P_{25}$ | $P_{50}$ | $P_{75}$ | $P_{95}$ |
|---|---|---|---|---|---|
| 1. Formal Assessment | 0 | 0 | 0 | 0 | 10 |
| 2. Informal Assessment | 0 | 0 | 0 | 0 | 1 |
| 3. Integrated Assessment and Instruction | 0 | 0 | 2 | 3 | 7 |
| 4. Other On-Task Actions | 0 | 4 | 6 | 6 | 8 |
| 5. Off-Task Actions | 0 | 1 | 1 | 2 | 4 |

Table 2: Distribution of Actions Occurring in the Middle School Classrooms

| *Actions of the Teacher* | $P_{05}$ | $P_{25}$ | $P_{50}$ | $P_{75}$ | $P_{95}$ |
|---|---|---|---|---|---|
| 1. Formal Assessment | 0 | 0 | 0 | 0 | 0 |
| 2. Informal Assessment | 0 | 1 | 2 | 5 | 8 |
| 3. Integrated Assessment and Instruction | 0 | 1 | 3 | 5 | 9 |
| 4. Other Actions | 0 | 0 | 3 | 6 | 8 |

| *Actions of the Students* | $P_{05}$ | $P_{25}$ | $P_{50}$ | $P_{75}$ | $P_{95}$ |
|---|---|---|---|---|---|
| 1. Formal Assessment | 0 | 0 | 0 | 0 | 7 |
| 2. Informal Assessment | 0 | 0 | 0 | 1 | 4 |
| 3. Integrated Assessment and Instruction | 0 | 1 | 2 | 4 | 9 |
| 4. Other On-Task Actions | 0 | 1 | 4 | 6 | 8 |
| 5. Off-Task Actions | 0 | 1 | 2 | 3 | 5 |

Table 3: Distribution of Actions Occurring in the High School Classrooms

| *Actions of the Teacher* | $P_{05}$ | $P_{25}$ | $P_{50}$ | $P_{75}$ | $P_{95}$ |
|---|---|---|---|---|---|
| 1. Formal Assessment | 0 | 0 | 0 | 0 | 10 |
| 2. Informal Assessment | 0 | 0 | 4 | 5 | 10 |
| 3. Integrated A · essment and Instruction | 0 | 0 | 0 | 3 | 8 |
| 4. Other Actions | 0 | 0 | 5 | 7 | 10 |

| *Actions of the Students* | $P_{05}$ | $P_{25}$ | $P_{50}$ | $P_{75}$ | $P_{95}$ |
|---|---|---|---|---|---|
| 1. Formal Assessment | 0 | 0 | 0 | 2 | 10 |
| 2. Informal Assessment | 0 | 0 | 0 | 1 | 5 |
| 3. Integrated Assessment and Instruction | 0 | 0 | 0 | 1 | 5 |
| 4. Other On-Task Actions | 0 | 0 | 6 | 8 | 9 |
| 5. Off-Task Actions | 0 | 0 | 1 | 3 | 8 |

The interpretation of Tables 1 through 3 is somewhat different for teacher versus students data. Using Table 3 to illustrate, row 2 under *Actions of the Teacher* indicates the amount of time high school teachers spent on informal assessment. Within this row, a value of 4 is listed at the 50th percentile point. An interpretation of this is that when the observed time of the six high school teachers are divided into 10-minute segments, half of these time segments included 4 or more minutes devoted to informal assessment, and half of the segments included 4 or less minutes devoted to informal assessment. Similarly, one-fourth of these time segments included 5 or more minutes devoted to informal assessment whereas three-fourths of the segments involved 5 or less minutes devoted to informal assessment. Likewise, in half of the time segments, 5 or more minutes was devoted to other actions, that is, activities that did not include assessment.

Unlike with teachers, observations of students simultaneously involved multiple individuals. The interpretation of this data is also illustrated using Table 3. Row 1 under *Actions of the Students* lists a 10 at the 95th percentile point. This indicates that when the observations of students at the high schools were divided into 10-minute segments, in at least 5 percent of the segments, all students spent the full 10 minutes involved with a formal assessment. A 2 is listed at the 75th percentile point. This means that one-fourth of these time segments included 2 or more minutes as the estimated *average* of student time associated with a formal assessment. An average of 2 minutes could mean that all the students were actively involved in a formal assessment for the first two minutes of the time segment, that 20% of the students were actively involved with the formal assessment throughout the ten-minute segment while others did something else, or some combination or variation of these events occurred.

(One might anticipate numbers at the 50th percentile to sum to 10, which they obviously do not. The reason they do not sum to 10 can be illustrated using a scenario involving an extreme situation. Assume that, for student observations, 20% of the time segments involved only formal assessment. For these time segments, a 10 would be posted for formal assessments and a zero for

each of the other five categories of activity. Further assume that 20% of the time segments involved only informal assessments, and similarly 20% of each of the remaining time segments were devoted fully to one of the other activity categories. The resulting matrix of activity categories by time segments would included 20% 10's and 80% zeros within each row. When sorted, a 10 would appear at the 95th percentile point within each category, whereas a zero would appear at all other percentile points listed in Tables 1 through 3, including at the 50th percentile point.)

Tables 1 through 3 indicate that, during observations, minimal class time was associated with formal assessment, whether these assessments be written tests, formally developed and administered performance assessments, or other formal techniques. These teachers spent substantially more time on informal assessments, such as observation and informal questions. If one includes integrated assessment and instruction, where overt assessment actions are integrated in instruction, within the observed elementary and middle school classrooms, assessment and assessment-associated activities accounted for the majority of the teacher's time. Substantially less than half of these teacher's time in the classroom involved activities separate from assessment.

Tables 1 through 3 likely underestimate the amount of time devoted to informal assessment. For instance, teachers continuously monitored student performance through observation, regardless of the activity the teacher was involved with. The existence of this continuous informal assessment became abundantly clear the moment a disruption occurred, and was also obvious in the moment by moment adjustments each teacher would make as activities evolved. A teacher's involvement with assessment was recorded only when the activity was obvert.

Relative to their teachers, the actions of students were more frequently associated with formal assessments but less with informal assessments. The majority of observed formal assessments involved written tests. As noted earlier, the teacher tends to be involved in monitoring students, an informal assessment, when students working on the test. In contrast, informal assessment activities of students more often involve one student at a time or small groups of students. When this subset of students is participating in an assessment-related action, other students are typically involved with something else. Nevertheless, particularly when integrated assessment and instruction is included, observed student activities more often were associated with non-formal than formal assessments. Teachers, however, appeared to spend a larger portion of time with informal assessments than did students.

The major role of informal assessment and integrated assessment and instruction that was observed in these classrooms is in stark contrast with emphasis provided in current measurement textbooks. Although the trend is to include more emphasis on informal assessment, the majority of texts concerned with classroom assessment limit discussion of informal assessment to a few paragraphs. In this regard, many textbooks seem non-responsive to assessment-related needs of teachers.

*Information Drawn from the Narrative Records*

During observations, narrative records were made describing student and teacher behaviors that occurred during the 10-minute time segments. The purpose of these records was to allow later reconstruction of the essence of what transpired during class. As noted earlier, emphasis was given to events or situations that appeared relevant to the set of questions that guided observations. A synopsis of the narrative records, as they relate to each group of these questions, is presented here.

A) What techniques do teachers use to assess students? How do these techniques compare to those discussed in textbooks on classroom assessment?

Informal assessment techniques clearly are dominate in the observed classrooms, particularly in the elementary and middle schools. These assessments depended largely on observations and oral questions. Here are some typical narrative records:

> The teacher asked for a show of hands of how many understood what she was talking about.
>
> Students have read written material concerning five types of propaganda. Teacher asked for volunteers to give examples of each type and called upon students who raised hands.
>
> Seatwork continues. Teacher walks around looking at student work.
>
> Student volunteers work (math) problems on board. Teacher asks by show of hands how many correctly answered each problem.
>
> Teacher described a science experiment. Teacher asked students, mostly through show of hands, to state conclusions that can be drawn from this experiment.
>
> Teacher reads aloud to students a short lesson as students read to themselves. Teacher frequently stops to ask questions about what was read; students who raise hand are called upon. Students then work in groups on exercise based on lesson. Five minutes in, teacher asks how far students are. Asks again two minutes later. Teacher then asks a member of each group to indicate what the exercise demonstrates.

Interestingly, when teachers were asked how they assessed their students, only formal assessment techniques were addressed. Perhaps, those of us who teach college courses would respond the same way. When asked about the many informal techniques they were observed to be using, each teacher was quick to recognize the important and dominant role informal assessments play.

As already noted, measurement textbooks that are aimed at classroom teachers similarly focus heavily or almost exclusively on formal assessments. In part, this emphasis may be a natural consequence of the background of textbook authors and college faculty who teach courses concerned with classroom assessment. We authors and college faculty come from and typically work within an environment dominated more by large-scale testing programs than by elementary and secondary school classrooms. However, we cannot justifiably plead ignorance. Highly visible writings by Bloom, Hastings, Madaus (1971) and Glaser and Nitko (1971), among others, clearly address the importance of integrating assessment and instruction and the role of formative evaluation to this integration. Authors such as Airasian (1994), Oosterhof (1994), among others, address the dominant role informal assessments play within formative evaluations. If the extensive amount of classroom time found during these observations to be devoted to informal assessment is at all representative, then there appears to be a critical need for including informal assessment as a more dominant part of courses and instructional materials associated with classroom assessment.

B) Do teachers' assessments appear to be valid? Does observed student performance appear to generalize to performance the teacher did not observe?

By their nature, the present observations do not allow a careful analysis of validity or generalizability of the classroom assessments. Some interesting indicators, however, appeared with consistency and seem worth addressing.

With respect to validity, one can use the conventional construct-, content-, and criterion-related categories of evidence to frame discussion. Regarding constructs, some teachers seemed more

adept at conceptualizing a construct, that is, establishing student performances that would provide a good indication of what students know or are thinking. This ability seemed related to the teacher's expertise with the content being taught. That expertise might be quite specialized. For example, one middle-school teacher was observed teaching science. This teacher's particular specialization is life sciences. When the lesson was related to the structure and functions of cells, the teacher's instruction, the questions asked of students, and related activities all appeared to be goal driven. The teacher seemed highly responsive to subtleties in student behavior and appeared to have clear ideas as to which student performances provide an indication of what students knew. In a separate lesson on astronomy, the teacher's instruction and assessment appeared to be more activity driven. Subtleties in a student's response seemed less useful to the teacher. A student's knowledge of a concept was more likely to be examined in terms of factual information rather than applications or implications of a concept. This pattern was consistent across teachers that were observed. When the teacher had a deeper understanding of the content, activities including assessment appeared to be more goal driven. When understanding was more shallow, the teacher's activities appeared to be more activity driven. A teacher's awareness of underlying constructs was less obvious when instruction was more activity driven. A statement by Stiggins (1991) is relevant here:

> One of the basic tenets of sound assessment in any context is that the assessor possess (a) a clear and highly differentiated vision or understanding of the achievement target to be attained by students and (b) a thorough understanding of the full range of assessment alternatives available to assess the target of interest (p. 8).

Although a teacher's knowledge of academic content is generally not the responsibility of a college course concerned with classroom assessment, this content knowledge appears relevant to the adequacy with which classroom assessment techniques are applied.

With respect to content-related evidence of validity, the observed teachers appeared to sometimes but not always collect this form of evidence. During conversation, the teachers indicated they did not use a table of specifications or a written list of objectives when developing a formal assessment. Some of the teachers simply used assessments provided with the curriculum materials without any evaluation of the appropriateness of its content. When assessment involved content with which the teacher appeared to have a more in-depth knowledge, the teacher seemed more likely to be uncomfortable with some or all of the assessment material provided with the curriculum. In this latter situation, teachers appeared to plan formal assessments by developing a mental outline of content that should be included and then developing the assessment from this conceptualization. From listening to these teachers' descriptions, one gets the impression that much the same content would have been established had a more formal procedure been used such as a table of specifications. It would be useful to establish through a more systematic analysis whether or not this impression is correct.

With respect to informal assessments, the content again appeared to be more goal driven when the teacher had a deeper understanding of the content. Informal questions asked of students were more typically created by the teacher. The teacher seemed more likely to adapt the content of oral questions in response to students' answers. When the teacher had a more shallow understanding of the subject matter, the teacher depended more heavily on exercises provided with the curriculum. Activities appeared to be more activity driven. In conversation, none of the teachers acknowledged planning the content of informal assessments. They said the content just happened, much like one conducts a causal conversation.

Teachers often *do* collect criterion-related evidence of validity. They, of course, do not use statistical correlations to establish relationships between test performance and an external criterion. They often do, however, correlated what was observed with other indications of a student's knowledge. A number of teachers were more tentative in their interpretation of student performance. For example, some teachers asked follow-up questions to substantiate a judgment, or cautiously interpreted an atypical performance on a quiz. Other teachers, in contrast, were more emphatic in their interpretation of a student's performance. At issue seemed to be whether or not the teacher recognized measurement error. Teachers who more tentatively interpreted a student's performance expressed a number of reasons for performance deviating from a student's actual ability. Reasons expressed included "the student may not have understood the question," "the teacher may not know why a student answered the way that he did," and "a student might not be concentrating." Teachers who more emphatically interpreted student performance tended to associate changes in observed performance with changes in the student. Statements such as "the student did not study," "the student knows this material better than other areas," and "each student finds it easier to learn some things than others" were typical teacher comments.

Validity issues related to the construct and content appeared to be linked to the teacher's understanding of material being taught, this varying across teachers, and within teachers across content. In contrast, the degree to which a teacher sought criterion-related evidence of validity appeared to be a function of whether the teacher anticipated possible discrepancies between observed student performance and actual student ability.

As with validity, the present observations do not lend themselves to estimating the generalizability of teacher's assessments. Some interesting patterns, however, did emerge, particularly with respect to informal assessments. Most students, even young students, appeared to have an uncanny ability to selectively avoid being called upon or observed. This was repeatedly observed by focusing on one student for period of time. Students would become visible by raising their hands, squirming, making noise, establishing eye contact, and through body language expressing excitement. Students would be less visible to the teacher by not doing these things. These attention-getting actions would be turned on a like a switch, possibly at the moment the student established what was thought to be a desirable response. Particularly among students who were older or more capable, some students would be selective as to how aggressively they would solicit the teacher's attention. A substantial number of students, particularly among students of lower ability in higher grades, would not participate, and typically were not called upon. In essence, informal assessments appeared to involve an unrepresentative sample of students. This would reduce the degree to which the informal assessments generalize.

During conversation, several teachers stated that practice teachers whom they had supervised often were very surprised with how poorly students did on their tests. This may be the result of informal assessments involving unrepresentative measures. Interestingly, some of the teachers who were observed indicated they expect students to do worse on a formal test than during class. One teacher acknowledge the phenomenon this way:

> Watching how students do during class, by itself, is not sufficient. Quizzes need to be used frequently as a reality check.

Some teachers appeared to be more effective than others in terms the use of informal assessments. Some are particularly careful to call upon students who are not actively participating, or to visit briefly with non-participants while the class is involved in seatwork.

C) What conditions and approaches appear most useful for helping teachers learn assessment techniques?

Most of the observed teachers had never enrolled in a course devoted to classroom assessment. They all had completed a teacher certification program and, sometimes bitterly, complained about the irrelevance of many of the education courses they had completed. In conversation, the teachers stated that many college faculty in education do not appear to know what is going on in the schools. Relating back to their own training and to the training provided practice teachers they have supervised, they believe the training of teachers tends to follow fads. Had these teachers completed an assessment course, they may have expressed these same concerns. Certainly, training in assessment, to be useful, must be responsive to the needs of teachers.

In assessment, one aspect of being responsive would be to recognize the important role informal assessment plays in the classroom. Issues such as gathering evidence of validity and determining whether what was observed generalizes to what was not observed should be carefully applied to informal as well as formal assessments. Another aspect of being responsive is focusing on skills teachers can apply within the classroom. Teachers can advantageously use the concept of reliability but have little if any need knowing how to compute a reliability coefficient. Teachers may benefit from knowing basic characteristics of standardized tests and being able to evaluate their common uses. Teachers seldom or never are asked to use the familiar references for critiquing or selecting a standardized test.

Among the observed teachers, there is considerable interest in alternative assessments. Part of this interest appears related to a genuine interest in more adequately assessing students. This interest, in part may be due to peer pressure. These teachers were unclear with respect to what alternative assessment involves. Perhaps in assessment we share some of that concern. Certainly, conversations with the fifteen teachers suggested a need to address portfolios, performance assessments, and authenticity as they related to classroom situations.

Within formal assessments, the observed middle and high school teachers placed most of the emphasis on traditional written tests, particularly those that use the short-answer and multiple-choice formats. Trends may be away from these formats, yet the need to help teachers become proficient at producing and scoring these test may still be significant.

Personally, one of the more surprising findings from the observations is the limited or non-existent time teachers have, when they are teaching, for developing assessment skills. Obviously, a college course in classroom assessment should place emphasis on the *application* of measurement skills. Significant amounts of time obviously should be devoted to actually using the principles and techniques that are taught. But giving a strong emphasis to application may be insufficient. After observing these teachers, one leaves with a distinct impression that teachers will not have the opportunity to advance assessment skills beyond the level of proficiency gained during training. If this is so, then there is a need to evaluate the effectiveness of instruction in assessment in terms of the skills our students have when they exit the course. If prospective or practicing teachers' abilities with critical measurement skills are less than acceptable, then perhaps we need to give careful consideration to selecting a subset of skills with which we will train teachers well.

## References


Airasian, P. W. (1994). *Classroom assessment* (2nd ed.). New York: McGraw-Hill.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw Hill.

Glaser, R. & Nitko, A. J. (1971). Measurement in learning and instruction. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 625-670). Washington, DC: American Council on Education.

Oosterhof, A. C. (1994). *Classroom applications of educational measurement* (2nd ed.). New York: Macmillan.

Stiggins, R. J. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice, 10*, 1, 7-12.