PRACTICING TEACHERS BELIEFS AND USES OF ASSESSMENT

Anjoo Sikka
Janice L. Nath
Myrna D. Cohen
University of Houston – Downtown
HOUSTON, TEXAS, U.S.A.

Abstract

In an exploratory study of teachers’ beliefs and use of assessment, four certified teachers of culturally and linguistically diverse students volunteered for interviews and submission of assessments they created. Detailed case studies, focusing on exploring the complexities of assessment-related issues, were developed. Common themes emerging from the case studies were: (1) a frustration with standardized testing, (2) low self-efficacy in preparing objective types of tests, and (3) concern about the value of objective test items for teaching. In the socio-political environment of high stakes testing, implications for testing practices, support for assessment, and teacher education and training are also presented.

KEY WORDS: Assessment, self-efficacy, case study, teachers

INTRODUCTION

Two decades ago, Bandura [1986] proposed that beliefs held by individuals direct many of their important decisions. Since then, an increasing amount of teacher research has sought to discover teachers’ beliefs and the impact that those beliefs have upon their classrooms [ISAS, 1996; Wilson, 1990]. This research is particularly critical for assessment issues, as Shavelson and Stern [in Chase, 1999] judged that “teachers make decisions requiring assessment information at a rate of once every two to three minutes” (p. 4). Given these findings, it is easy to see that assessment-related activities have been found to occupy at least one third to one half of a teacher’s time [Stiggins & Conklin, 1992]. Brookhart [1998, as cited in Mertler, 2004] underscores this idea by pointing out that classroom assessment is, in essence, connected to every other aspect of teaching and informs instruction.

Although teachers often rely on day-to-day observation and information to make decisions [Stiggins, 1994], assessment knowledge of informal and formal procedures guarantees a fairer and more just evaluation. As demand increases for more concrete evidence for justifying judgments about students’ work, placing students in various programs, receiving funds for student achievement, and so forth, many educators, parents, and state and national governments have become more interested in what teachers know and believe about assessment. Mertler [2004] points out that, although expectations for teachers’ assessment skills have risen, many teacher preparation programs do not require preservice teachers to take courses in classroom assessment, and inservice teachers report that they were not well prepared to assess student learning. Research has, in fact, found that teachers are neither well prepared in their knowledge of classroom assessment nor in large-scale testing.

Mertler [2004] questions whether assessment training for teachers is more effective for preservice teachers, or for inservice teachers who would learn about it as “on the job training” where it can be contextualized. In his study he found that preservice teachers knew much less about assessment than their inservice counterparts. This was especially apparent among secondary teachers. He concludes that
further investigation is needed because, “The ability to assess student performance in appropriate, valid, and reliable ways is arguably one of the most important aspects of the job of teaching” [p. 63]. It is also widely accepted that poor assessment instruments can often be designed and/or used, and those who are ill-trained or naïve can misuse even the best of assessment tools [Worthen, Write, Fan, & Sudweeks, 1998]. Ward and Murry-Ward [1999], for instance, note that “lack of knowledge makes (teachers) uninformed users of assessments, so they do not even know the critical questions they should ask about the instruments they use” [p. 9]. Ward and Murry-Ward continue to acknowledge that there is a lack of training in educational programs in assessment and that the training that teachers may receive may not be what teachers want or need.

One critical issue that has begun to affect teachers’ beliefs about assessment in past years is high stakes testing (i.e., situations where test results have a significant impact on the lives of children, ratings of schools and its personnel, funding, etc.). Although educators have worked to move the focus of assessment from memorization to more authentic forms of assessment [Gredler, 1999], high stakes accountability continues to alter instructional focus (narrowing curriculum and emphasizing teaching of test-taking skills). For example, Abrams, Pedulla, and Madaus [2003] found that increased attention by teachers on content that will be tested has led to reduced emphasis on those curricular areas that will not be tested. Moreover, high stakes testing can result in teachers implementing strategies and practices that go against their beliefs about learning and best practice (Abrams, Pedulla, & Madaus, 2003). High stakes testing was also found to decrease teacher morale, increase teacher attrition, and contribute to the de-professionalization of teachers.

In recognition of the importance of assessment skills among teachers, secondary teachers in Texas are expected to master thirteen competencies for their state certification exam (TExES). One of the thirteen competencies, Competency 10, directly addresses assessment. Preservice teachers are expected to know about various assessment methods and their advantages and disadvantages. They are also expected to be able to create valid assessments, provide appropriate feedback, encourage student self-assessment, and adjust instruction according to ongoing assessment of student performance. Assessment concepts are tangentially noted in other competencies too, such as in Competency 2 (which addresses diversity and assessing students of diverse backgrounds) and Competency 3 (which addresses planning and assessing lesson effectiveness as well as understanding the Texas statewide assessment program). It is logical to expect that these state teacher competencies are part of teacher preparation programs in Texas and that their prominence ensures that teachers begin their careers with some fundamental knowledge about assessment in Texas. In addition, the State Board of Education does rely on high stakes standardized testing of students to make decisions about student promotion to the next grade, along with teacher and school performance. Student pass rates for schools feature heavily in local and state media, even to the point of influencing the real estate prices in the “catchment” area. Hence, Texas teachers often have to balance a variety of demands on their time, including that of standardized test preparation. This reliance on standardized test scores has now spread to various parts of the United States.

What does this environment mean for a teacher’s everyday planning, teaching, and assessment activities? How do teachers perceive the role of assessment in this context? There is a significant need to examine assessment-related beliefs and attitudes, on the one hand, and effective use, on the other hand, using a wide-angle lens. Assessment-related decisions that teachers make are both public and private and are influenced heavily by both political and personal factors. In order to tap this complexity, we use case study methodology to explore teachers’ beliefs about and use of assessment.

**METHODOLOGY**

The purpose of this study was to examine secondary teachers’ beliefs about, and practices in assessment using the case study methodology. The case study approach, which is qualitative in nature, has a very important role in examining the complexities of phenomena, particularly for exploratory research. For instance, Darling-Hammond [2006] has used interviews extensively to explore teacher candidate views and beliefs.
PARTICIPANTS

Four full-time secondary school teachers, who were enrolled in courses for either an Alternative Certification Program or a Masters of Arts in Teaching degree, were recruited to participate in this study. Researchers asked permission of university instructors to take a few minutes of class time to explain the purpose and needs of the study and to distribute a written letter of invitation to eligible students. Two researchers were teaching eligible participants, so only non-instructor researchers recruited these students in order to avoid feelings of coercion. After acquiring teacher volunteers, researchers made decisions regarding division of interview tasks, taking care to avoid assigning participants to researchers who had previously taught them at the university.

Participants’ years of teaching experience varied from 1-8 years. The content areas these teachers (as a group) reported teaching were Science and English/Language Arts. All four participants are female, with one student’s ethnicity being White and the other three being Hispanic. Participant age ranged from 24 to 34 years. Most teachers had received some training (a course) in assessment. In all cases, the students these teachers taught are from culturally and/or linguistically diverse backgrounds. In addition, they also reported teaching students with disabilities.

PROCEDURE

Researchers informed the volunteers that they were to participate in an hour-long interview to which they would bring (1) copies of assessments that they had used in the past month and (2) a description of a brief assessment-related case from their classroom (with either successful or unsuccessful results). In three out of the four cases, teachers did not provide a description of the assessment-related case. Each participant received a monetary reward ($25.00) for their participation.

Each interview session began with a short paper/pencil demographics questionnaire. The interviews were audio taped and transcribed. The interview protocol included questions about the teachers’ perceptions and uses of assessment in the context of their schools and districts, and their perceived competence in creating and analyzing assessments. Questions about how assessment of special populations such as second language students and students with disabilities were also included when relevant.

Data were analyzed using manifest and latent content analysis of responses to the interview questions and the assessment artifacts submitted by teachers. Analyses were focused on the knowledge, beliefs, concerns, benefits, and training experiences of each participant. Each case was summarized in a two- to three-page case study organized around common themes and then subjected to a cross-case analysis for emergent themes. The analysis was conducted by a single researcher, with the other two researchers providing validation of general inferences and conclusions.

RESULTS

Case studies were developed on the basis of responses to interview questions which were then checked against the issues that emerged from an examination of the assessment artifacts submitted by teachers. The purpose of this approach within the case study methodology was to capture the complexities of beliefs about, perceived use of, and knowledge of assessment.

SARAH

Sarah is a middle-school reading and language arts teacher who works for a religious school. She holds an undergraduate degree and has completed requirements for certification. She is in her mid-30s, Hispanic, and teaches children who are primarily Asian-American (94%). She has not taken a course in assessment and has been teaching for 8 years.

General themes

In her interview, Sarah expressed her frustration with balancing the pressing need for accountability (and associated mandated testing) with the needs of her students. She reported grappling with the conflict of making teaching and assessment relevant, authentic, and challenging.
interview, this frustration was only seconded by the challenge of motivating students or, more precisely, preventing students being “turned off” academics. Time, technology, and space limitations were other areas of frustration for Sarah.

Another concern that she expressed was the need to mediate between (decisions made by) administrators and students. She felt that she had to work with administrators to mitigate the effects of decisions made regarding assessment that might cause pressure or be de-motivating for students. She expressed the belief that practices like “standard” assessments and pre-testing for knowledge/skills that had not yet been learned were stressful for students and may well have the effect of creating or substantiating negative self-perceptions on students’ part.

**Use of Assessment**

A theme that emerged from her responses to various questions was that teaching is an integral part of assessment. She felt that she spent a considerable amount of her time (33-50%) assessing students (formally and informally) and relied on assessments to make decisions about student learning and her effectiveness as a teacher.

“In the classroom, what I do is listen...I give them prompts to restate the questions that I ask. I do a lot of oral testing. Later, I have them work in groups and ask them to assess each other (peer assessment). I allow them to do presentations where they are allowed to do more creative stuff. It’s anywhere between one-third to half of the time”

Although she realized that assessment has many uses, she rated the purposes of identifying student strengths and weaknesses and re-teaching as most important, followed by the purpose of determining if students have learned and providing them and their parents’ feedback. Contrary to the popular assumption that has guided the emphasis on assessment, she did not believe that assessments are very useful in motivating students, giving it a rating of “3” or “moderately important” on a 5-point scale. In addition, she did not feel that the self-evaluative (teacher) function of assessment was even moderately important.

As far as her preference for assessment was concerned, she showed a tendency to use projects, presentations, and essays. She reported that, in the most recent year, she had incorporated multiple choice questions in her repertoire, mostly because it was required by her school’s administration, but also because she had been exposed to teacher certification exams (which were multiple choice) and realized that this format can also be used to create challenging tests.

She preferred to use a game-like format for assessment that aimed to determine if the students are learning. She showed a considerable amount of empathy for students, emphasizing that it was very important for students to enjoy their learning.

In providing feedback, she showed a preference for giving detailed, annotated feedback, much in keeping with her English/Language Arts background. She reiterated that student learning and engagement were paramount.

She did not report involving parents in assessments, noting one exception when she asked students to interview parents in a “family tree” activity. Her interactions with parents about assessment mostly pertained to informing them about the performance of their child, with a school-mandated call to parents if the student failed a test.

**Areas of weakness, need for future learning**

Sarah expressed some concern for addressing the needs of Second Language Learners and students with disabilities. She felt that support in these areas was not provided by the school or district. In addition, she felt that she needed help in developing multiple choice items, portfolios, and projects over extended time. She felt she also needed help with grading observation checklists and laboratory activities. A primary area of concern that emerged was preparing tests that were free of bias, that had high validity (i.e., assessed the objective(s) for which the test was intended) in general, and for Second Language learners in particular. Sarah reported using “rubrics” to assess student work, but seemed to rely mostly on those developed by others. Although it might be presumptuous to believe that she may not know how to develop rubrics, such lack of knowledge/skill in developing rubrics is common among trained teachers also. Sarah also expressed an interest in learning how to organize data from assessments to make decisions about students’ learning.
Affective and evaluative responses to practices in assessment

Implying that there is considerable invalid use of assessment, Sarah reported that standardized tests have a negative impact on students. In addition, she seemed to disapprove of the use of test data to compare schools and teachers. She felt that, given that teachers have little control over student learning (beyond the classroom), evaluating teachers on the basis of standardized assessment was inappropriate. She preferred to tailor the assessment to student needs.

Sarah also showed a preference for teachers making decisions about the best way to assess students. Even though she complies with the testing requirements in her school, she showed some dissonance between her belief and practices and spent a considerable amount of time incorporating both approaches to assessment in her classes. Of course, she also reported a shortage of time for teaching.

Her reported reason for the popularity of multiple choice exams was ease of grading. She did not seem convinced, at the point this question was posed, that multiple choice tests serve a purpose of authentically assessing learning.

Triangulation of teacher self-reports with submitted assessments

Sarah submitted copies of assessment that she had used in the recent past. These consisted of a mix of objective and essay questions, and a copy of a rubric that she had obtained from the State Education Agency. Her high self-efficacy with essay tests and judging performances were substantiated by the artifacts she provided in this area. The assessments measured high level of thinking, were challenging, and directions for taking these tests or fulfilling the performances were thorough. There was one error in a supply-type question (a key-type multiple choice section -- where the options to 7 or 8 multiple choice questions are the same). The Unit tests that she referred to in her interview and which are mandated by the school administration were multiple choice and alternative choice (true-false or some such dichotomy). The questions assessed lower levels of thinking -- primarily knowledge, comprehension and application levels. In addition, multiple choice questions had several errors. Some examples were: (a) the stem did not pose the question and was very open-ended; (b) most multiple choice questions had only 3 options, rendering them just a little better than alternative-choice; (c) the format of multiple choice questions was not predictable--in some cases, the options were stacked and in others, they were presented in one line; (d) directions for taking tests were not detailed, instead they seemed to be standard and short.

Conclusions

In general, Sarah’s frustrations with the conflicting pressures of assessment, her need to take care of her students’ learning, and her cynicism about assessment were self-evident. She seemed open to learning new types of assessment but sees herself as the primary force in her students’ learning. The teaching-evaluative and comparative functions of assessment (current political climate in education) were rejected by her.

MARGARITA

Margarita is a 26-year-old Hispanic teacher who has been teaching for three years. She is a certified teacher, currently pursuing a masters degree in teaching. She teaches English, Reading, and advanced English/language arts at a middle school in a suburb of Houston. Her student population is fairly diverse with 50 percent being African-American, 35 percent Hispanic, 10 percent White, and 5 percent Asian. Out of the 90 students whom she taught that semester, 15 students are English Language Learners, and three students have disabilities. She has taken a course in assessment.

General themes

Margarita showed considerable curiosity about testing procedures. She uses assessment both to provide feedback to students and to inform her teaching. She works in a setting where classroom teachers rely considerably on each other. Even though instructional specialists may be available, Margarita reports not using their assistance because of issues with time and scheduling. She expressed some degree of frustration with standardized tests but seemed to be able to incorporate their use into her repertoire of assessment approaches.
She reported a high level of efficacy with both extended response assessments (homework, projects, demonstrations) as well as simple response assessments (multiple choice, matching).

“When I arrived at middle school [from elementary school] though, I learned about the Scantron machines. And I fell in love with it (sic). We just don’t do Scantron stuff too much in elementary.”

She did, however, express concern that multiple choice tests do not really provide useful information.

“Again, it’s not anything that allows true honest feedback of the ability of that particular person.”

She seemed to have mixed feelings about multiple choice items. On the one hand, Margarita felt very comfortable using them and uses them in all major tests (but not quizzes) with enthusiasm but, on the other hand, she emphasizes the reason they are used is related to convenience and other practical considerations and that they do not provide “real” feedback.

Margarita also showed a high level of concern for students in almost all the questions she was asked. She provides detailed feedback, sometimes in personal consultations settings. She seemed to take a very cognitive approach to improving student motivation and learning (i.e., explaining why they made the score(s) they did and providing considerable corrective feedback).

**Use of Assessment**

Margarita rated a variety of “purposes of assessment” as highly important. These included: motivating students, assessing how to teach differently, identifying student strengths and weaknesses, placing students in special programs, informing students about their own learning strategy, informing students about the extent to which they have mastered the material, and communicating student progress with parents. Only two other purposes of assessment were rated as less important (but still a rating of 4) – evaluating herself and seeing how students are performing in relation to others.

In terms of her reported use, Margarita reported using a variety of different approaches: presentations, projects, essays, vocabulary tests (1 -2 word answers), and for larger tests, multiple choice questions. She reports spending about six hours every week on assessment-related activities, while her students spend about two hours per week. In her interview, she also admitted to using another approach, reciprocal teaching, to assess students. She seems to use a variety of assessments – both formal and informal. In fact, when asked to provide a definition of assessment, she reported it as “anything that allows the teacher to determine if the student has learned”.

In keeping with her language arts background, Margarita did show a preference for extended responses assessments, but when asked directly about the most commonly used assessment in her class, she mentioned vocabulary quizzes. She linked the emphasis on quizzes to student background, noting,

“...we use quizzes a lot - vocabulary quizzes – because we want to make sure [when] they leave eighth grade, they’ve acquired a little more vocabulary than what they may not have learned before, so we want to make sure they are okay with the words, especially because we have a lot (or at least I have a lot) of students of ethnic background (sic), and they don’t hear that kind of academic vocabulary at home...”

This quote also substantiates Margarita’s belief that assessment is important in a variety of ways, including checking for learning and for motivating students.

In providing feedback, she showed a preference for giving detailed feedback, using rubrics, and scheduling individual conferences with students. An examination of her assessment artifacts also showed the use of a wide variety of formal approaches – rubrics, vocabulary tests (92 questions chunked into sets of 3 to 5 vocabulary “matching” tasks), and multiple choice tests. She also reported using the rubric as a communication tool, handing out a copy of the grading rubric to students before they developed their writing task, which is noted as best practice with the use of rubrics. Other approaches reported were using reciprocal teaching, projects, and presentations.

She did not report involving parents in assessments, in general, but stated that parents were involved in assignments. She also indicated that she had more interactions with parents of “level” (i.e., on-level) students than the pre-Advanced Placement class. She mentioned that the parents are well informed by her:

“But the parents are pretty involved, they are pretty aware for both groups in term of assessment, but [for] assignments – one group is more involved than the other simply because that’s what they know how to [be]...parents who are on top of their children having to strive for excellence at all times.”
Her reported reason for including multiple choice tests in her classes was to provide students with opportunities to practice for the state-mandated test (Texas Assessment of Knowledge & Skills – TAKS). Eighth grade (one of the two grades Margarita teaches) is a critical year for TAKS performance.

**Areas of weakness, need for future learning**

Margarita showed a high level of efficacy and enthusiasm for a variety of assessment approaches, equating multiple choice, paper-pencil tests with more “traditional” type of assessment. She reported lower efficacy in areas such as preparing research paper assessments, portfolios, and anecdotal records. In terms of grading, she also reported lower efficacy in assessing research papers, portfolios, anecdotal records, and observation checklists.

Portfolios and anecdotal records were mentioned again, when Margarita was asked to name the areas in which she wanted to learn more. When asked specifically about groups of students, she also reported that modifications that she made for students with disabilities were really not necessary, except in one case. She indicated that some students may use disability as a crutch and would perform well on tests prepared for non-disabled children.

In assessing children who are English Language Learners, she expressed concern that students take much longer to learn the academic language than school policies recognize. She expressed some frustration that such students may not be ready to take the required standardized tests.

**Affective and evaluative responses to practices in assessment**

Margarita recognized the heavy assessment emphasis in her school and classroom. She seemed to express some negative affect regarding this but participated herself in this assessment culture. Even though she reported that students spend one hour per week on assessment, it is likely that more time is actually spent on this. She uses assessment as a way to monitor learning, although she really did not mention how she uses assessment scores to evaluate herself. She expressed frustration regarding student performance (i.e., if they do not perform well on tests) and reiterated that tests may not tap into what students actually know.

**Triangulation of teacher self-reports with submitted assessments.**

Margarita submitted a large number of assessment artifacts. Her tests are fairly long, including approximately 90 vocabulary items in one instance. She uses a variety of approaches: essays (both extended response and restricted response), vocabulary quizzes in matching format, scoring rubrics, and analytic criteria for grading student essays. She does not use portfolios, observations, or anecdotal records. Test guidelines seem to be well constructed but are long, thorough, and fairly traditional in nature. Rubrics are also clearly used as a communication device.

**Conclusions**

This teacher has a high level of efficacy with using a variety of assessment approaches but felt that she needed support in using portfolios and anecdotal records. Her use of assessment includes both authentic and paper-pencil tests, with the artifacts being more reflective of a high use of multiple choice/matching-type questions. It is possible that this emphasis is a result of the perceived need to provide students with opportunities to practice with the standardized exam for which teachers and students are held responsible at the end of eighth grade.

**REBECCA**

Rebecca is a 24-year-old white female teacher pursuing a masters degree in teaching. She teaches at a special school that aims to prepare minority (97% African-American) students for college. She is a certified teacher and has been teaching for one year. She has taken one assessment course. She teaches ninth grade English. In her school, students are required to take a standardized test during tenth grade to monitor their learning and progress.

**General themes**

Rebecca showed a lot of empathy for her students and repeatedly expressed concerns about making the test relevant for her students. An examination of her testing artifacts substantiates this concern. She
uses examples of “Black” English, which students then translate into Standard English. The books she uses are appealing to this age and ethnic group.

She showed some degree of dissatisfaction as well as low self-efficacy in preparing multiple choice (or objective) tests. This was apparent from her responses to questions regarding areas that she would like to improve. However, she rated herself very high on ability to prepare multiple choice tests. An examination of the tests she has recently used in her classroom suggested moderate to high level of skill at preparing multiple choice items. However, some errors and clues were evident.

She seemed to be very aware of her students’ response to, and frustration with, mandated assessments. In her description of a case pertaining to assessment, she expressed her frustration with tests being used as “fillers” for some available time. Recalling one such incident, she noted that many students either did not attempt the essay questions or expressed concern that they were not able to finish the test in a designated time and that they worry about the effect of this on their grades.

“We created two enormously long essay questions they had to write. We put in as much stuff as we could cram into it and were feeling successful because it was long enough. However, I knew that my students would hate it and that half of them would think it was pointless and long and simply refuse to do it. They know that this is not a writing class and to them it felt as if they were being given an English exam.”

Her concern with relevance was reiterated in this scenario. She concludes, “All in all, it was a miserable two hours for all involved. Because of the lack of thoughtful preparation and unfair motives on our part, the kids did poorly on the exam and were frustrated with the process, whether they chose to actually complete the exam or not.”

Use of Assessment

Rebecca felt that assessment is useful when it is authentic, allows for student creativity, and provides an idea about student learning. She does not believe that multiple choice tests provide such information and, in fact, felt that these exams may inflate grades.

She realized that assessment is most important in informing teaching strategies, letting students know the extent to which they have mastered the material, in communicating progress to parents, and in evaluating herself. Purposes considered moderately important were giving students feedback on their strengths and weaknesses and their own learning strategies. Rebecca felt that the role of assessment in motivating students, placing them in special programs, and determining student performance in relation to others was slightly important.

As far as her preference for assessment was concerned, she showed a tendency to use essays, papers, writing assignments and, less frequently, presentations. She reported that, in the most recent year, she had incorporated multiple choice questions in her repertoire, mostly because it was required by administration and she wanted her students to be comfortable with these kinds of tests.

In providing feedback, she showed a preference for giving detailed feedback and using rubrics. An examination of her rubric revealed that she had 22 elements in that rubric (much more than most assessment experts would advocate).

She did not report involving parents in assessments, noting one exception when she asked students to interview parents for a “Life in the 60s” paper.

Areas of weakness, need for future learning

Rebecca felt fairly comfortable with assessing students with disabilities – indicating that she might, at times, reduce the number of questions or make the wording simpler. She felt that she might benefit from more assistance in preparing multiple choice tests, using rubrics, and ruling out biasing factors such as poverty. She expressed concern about validity of the tests – i.e., how can tests be made to measure what they are meant to measure. Although she did not mention these as an area where she wants to learn more, she did indicate a low ability in preparing portfolios and assessment over extended time. In her need for future learning, she emphasized tests (implying multiple choice questions) and making tests more valid (and free from bias due to students’ lower economic background).

Affective and evaluative responses to practices in assessment

Rebecca expressed frustration about three major issues: (a) overemphasis and misuse of tests; (b) concern about tests’ impact on her students; and (c) mandated assessment that does not really give a
valid picture of student learning. She is a relatively calm person but showed some discomfort when it came to the emphasis on standardized test scores and multiple choice format. It can be inferred that she believes that these practices are more a hindrance to learning than providing valuable information about students’ learning.

She expressed the belief that she has better knowledge about whether students are learning than these tests can reveal. She also recognized the importance of test scores in obtaining and sustaining funding for the school.

Her reported reason for the popularity of multiple choice exams was higher grades and a more direct connection to numerical scores. She did not seem convinced that multiple choice served a purpose of authentically assessing learning.

Triangulation of teacher self-reports with submitted assessments
Rebecca submitted copies of several types of assessments. An examination of these artifacts reveals some patterns. She did rely heavily on supply-type or written questions. However, the tests were not well organized, where students had to move from multiple choice to essay questions multiple times in the same test. Essay questions were at the comprehension, application, evaluation, and synthesis levels. However, multiple choice questions seemed to be more at the knowledge level, with about 20 – 30 percent being at the evaluative level, requiring some degree of inference regarding characters. However, it is difficult to make decisions regarding the cognitive level of objectives assessed without the context instruction (e.g., copies of the lesson plan).

Conclusions
This teacher showed a high level of sympathy and concern for her students and felt that she might need to protect her students from invalid and unfair assessments, particularly tests used as fillers or practice for standardized tests. She expressed a need to improve her ability for creating multiple choice tests and grading with rubrics.

KATARINA

Katarina is a 30-year old, Hispanic, certified high school science teacher who works in an alternative school for students who need accelerated or other altered learning environments. She teaches courses in biology, physics, environmental science, and teen leadership. She teaches 90 students, with 55% Hispanic, 30% African-American, and 15% White students. She has been teaching for three years and has not taken a course in assessment.

General themes
Katarina is the only science teacher who participated in this research study. Her concerns seemed to be more in the area of creating authentic learning and motivating assessment for her students. It is also worth noting that 90 percent of her students are parents (88% female). Themes that emerged out of the interview were frustrations with balancing the need for preparing students for the standardized assessment and that for assessing learning. She reiterated several times that standardized tests (specifically, multiple choice items) were not a good measure of learning and, at times, were unfair to students (i.e., they do not reflect what the students know). She was also concerned about motivating students to learn and some degree of student preoccupation with test performance.

It could be inferred that Katarina sees herself as a protector of her students against high stakes accountability. She reports using a variety of approaches alternative to multiple choice assessment, like peer teaching, oral questioning, reciprocal teaching, portfolios, etc. She repeatedly expressed more faith in these approaches to assessment than the more traditional forms of assessment.

Katarina views assessment as being integrated with teaching, where the two activities could actually be indistinguishable from each other. She had specific concerns about high stakes accountability testing, particularly for the individual students and for teachers.

Use of Assessment
Besides the clearly evident belief that teaching and assessment are integrated, Katarina reported spending a considerable amount of time (six hours per week in class with students and 15 hours per
week out of class) on assessment activities. However, she repeatedly noted the scarcity of time and expressed a sense of competition between time spent on assessment and "real" teaching. Her school uses departmental exams, developed by experienced teachers, for each unit. In addition, she uses teacher-made quizzes, projects, oral questions, and group projects. All students are required to pass the Texas Assessment of Knowledge and Skills (TAKS) to graduate. In her ratings of the different purposes of assessment, she recognized that all interviewer-provided purposes were very important or important. She gave six purposes the highest rating: assessing how to teach differently, identifying strengths and weaknesses of students, placing students in special programs, informing students about their own learning strategies, communicating progress to parents, and evaluating herself. Purposes that she gave a lower rating (but still considered important) were: motivating students, informing them about the extent to which they have learned the material, and seeing how students are performing in relation to others. In general, it can be inferred that she does assign a high value to assessment but is concerned about its impact because of the high stakes accountability associated with it. She reported using teacher-made quizzes and oral questions frequently in class. She also regularly uses the unit exam (standard for all teachers). She showed a considerable degree of enthusiasm for using oral questioning, projects, and peer instruction but expressed concern about amount of time that these required and the need to monitor how students performed on tests with formats similar to the standardized exams. She reported providing students with individual feedback and guidance on test-taking strategies. In providing feedback, she showed a preference for giving verbal feedback on the mental processes they used to answer questions (e.g., in calculations, or in multiple choice exams, determining the correct answers) in addition to feedback about student learning of concepts. She involves parents in assessments, usually requesting parents to provide students with the support necessary to complete an assignment and following up with parents whose children make unacceptable or low grades. She did express time constraints as a limiting factor in her contact with parents.

Areas of weakness, need for future learning

Katarina expressed a high level of concern about the validity of multiple choice exams, rating her ability to prepare these exams as "3" (average) on a 5-point scale. She also assigned an "average" rating to her ability to develop extended response essays, homework, and matching test items. In terms of grading, she expressed a lower self-efficacy regarding projects that involve extended time, portfolios, anecdotal records, and observation checklists. She reported using rubrics but obtained them from a well-known web site. She reported using alternative approaches to assessing students with disabilities but expressed concern that these students have to get used to the standardized test formats because this is the tool used to determine eligibility for graduation.

Affective and evaluative responses to practices in assessment

Katarina expressed a highly negative attitude toward standardized tests (inferring the use of multiple choice questions) and use of these scores to determine if students could move on to the next grade and graduate. She also expressed a conflict between assessing for performance on standardized tests and assessing for learning and teaching. She felt these were two separate assessment activities. In addition, she expressed a dichotomy between the multiple choice test and instruction in authentic settings, as exemplified by her statement:

"Why does it have to be multiple choice? Why can’t we take them out in the field and let them see what a caterpillar is. Take them to the medical center and see what a cell actually looks like. Why must we stay focused on one thing and not open their horizons?"

It is puzzling that she has created a dichotomy between multiple choice questions and authentic instruction. However, the researchers believe that this perceived competition between two activities is best explained in terms of scarcity of time and an expectation that assessment served an instructional purpose. In addition her comments may stem from a negative attitude toward mandated multiple choice testing.

Her concerns about this type of testing are also evident in the following comment:
"They are easy, they are quick, put them in a Scantron machine, and they’re graded. That’s it. I think that’s why they are popular. I think I’ve heard some people say it gives them the opportunity of [using
the process of elimination. Well yes, but it doesn't really stick, from my experience. If they [the teachers] let them [the students] teach the subject, then the kids are more likely to actually stick to it and say "Aaah, I got it," instead of the multiple choice a, b, c, d."

She also expressed negative beliefs about standardized testing, primarily because of her assertion that test scores are not reflective of what students know.

"I am so against...standardized test. Because I have some excellent students, they are bright but they just can't pass either one of the tests because of society, and because "what have you"—and they can't graduate. I have some students that...have told me 'I am not college material'. They don't graduate and they don't move on in life because they cannot pass the test."

As can be seen from the quote above Katarina probably believes that these tests may be a deterrent for some students, negatively impacting their likelihood of pursuing post-secondary negative outcomes.

In opposition to these negative attitudes, however, Katarina finds assessment to be satisfying in its function of validating student learning and enhancing student self-esteem, as exemplified by the following comment:

"We [her school] have labels of bad kid's school, and when our kids do better than the other high schools then it's like, 'Yeah, see? We're not the bad kids or the stupid kids'. That happened last semester. My kids scored higher than the two high schools."

Triangulation of teacher self-reports with submitted assessments

Katarina submitted copies of the final exams she developed for use in her class. These questions are primarily multiple choice (assessing knowledge, comprehension, and application). These include both traditional and interpretive multiple choice questions. She reported using test banks but stated that she has discontinued using them in the past year in favor of school-mandated unit tests. An examination of the questions she developed revealed that her self-efficacy related to creation of multiple choice tests is lower than demonstrated by the items she created. With the exception of some common errors (like using "all of the above" option only in questions when it is the correct option and, hence, giving clues to students) the quality of the items and test is high. In addition, she uses a two-column format for multiple choice options in favor of the easier, "stacked" approach, most likely to minimize consumption of paper.

She provided copies of multiple choice exams only, limiting any triangulation to this area.

Conclusions

Katarina showed a very high concern for the learning needs and preferences of her students and expressed a negative perception of multiple choice tests and the high stakes accountability associated with standardized tests. She spent considerable time on assessment activities, and it can be inferred that she believes that information from assessment is a valuable determinant of learning and teaching.

DISCUSSION & IMPLICATIONS

GENERAL FINDINGS

The primary finding of this study was the concern expressed by all teachers regarding the pressures of assessment, as mandated by high stakes assessment requirements of schools. Their school administrators emphasized standardized test practice tests and a specific format of tests (mostly multiple choice) and held them accountable for student performance on these types of assessment. Other studies have agreed with this perceived pressure of high stakes assessments in increasing stress and decreasing morale among teachers [Abrams, Pedulla, and Madaus, 2003]. In their study, these researchers reviewed other research, noting that more than 77 percent of teachers surveyed indicated decreases in morale, and in another study, 76 percent reported that teaching was more stressful since the North Carolina state-testing program had begun [Jones, 1999]. Over half of Maryland teachers and about 75 percent of Kentucky educators who participated in a study indicated that morale had declined as a result of the state test [Koretz et al., 1996a; Koretz et al., 1996b], and 85 percent of Texas teachers surveyed by Hoffman, Assaf, and Paris [2001] agreed with the statement that "some of the best teachers are leaving the field because of the TAAS [a high stakes state test for school children given at various grade levels]."
In the current study, teachers expressed frustration resulting from the cognitive dissonance that arises from having to use a multiple choice format for testing when they believed that other types of assessment (essay, performances) were more valid.

“I’d love to have more open-ended questions, more cooperative learning,” noted one participant. Another noted:

“I have a couple of students who are just horrible, horrible test takers. You know, there’s just not anything they can do about the ‘bubbling thing’, and they are just awful test takers but they are very creative in other ways. And both are very, very different students, and I see that they have acquired the same skills, but they just need to be assessed a little bit different…through performances, through story telling narrations, through independent work, through feedback, meaning just question/answer kinds of thing. Sometimes I even allow the students to teach. I tell them that I’m tired, and they need to take over…and I’ll guide them through the lesson, and they’ll help teach the class. And they feel important when they do that, and it makes them more responsible in the terms of -- oh my gosh, I have to know this in order to teach the class.”

This matched research by Abrams, Pedulla, and Madaus [2003] who found concern about these types of test scores being used as a measure of student achievement.

In this study, teachers expressed concern about the performance pressures placed on students and teachers in using multiple choice, standardized test scores. Two participants also expressed grave concern about the use of test items that were beyond the level of students and the resultant negative impact on students’ self-esteem and motivation. Haney [2000] and Reardon [1996] have also associated the use of high stakes testing with an increase in the rate of student dropouts.

Teachers had extensive knowledge about assessment, with most teachers being familiar with the gamut of assessments that can be used in the classroom; however, they tended to use the more traditional forms of assessment in their classes for the purpose of assigning grades. As far as usage of different types of tests was concerned, teachers used a wide variety of assessment approaches. It could be inferred that teachers used traditional tests for the purposes of: (a) familiarizing students with this format, since it is used for standardized assessment; (b) providing students feedback in this context, hence training them in the thinking processes involved in answering questions in standardized tests; (c) complying with department, school, or school district requirements; and (d) communicating data about student achievement to external groups like parents, principals, and others. Alternative approaches (such as peer teaching, projects, oral questioning) were more likely to be used for informing teaching and learning. Teachers reported using rubrics to provide students with feedback on projects or other extensive assignments. All teachers reported having disagreements with their colleagues in the area of assessment, usually relating to length and/or type of assessments.

Three out of four teachers reported not involving parents in assessment, with one occasionally assigning projects that involved them (e.g., interviewing parents). In three out of the four cases, support for assessment activities was low and tended to be incorporated in general support pertaining to instructional support. However, this did not deter teachers from seeking knowledge and input from colleagues. As mentioned above, these consultations were likely to result in disagreements but tended to be handled more as discussions, as exemplified by the comment below:

“We don’t toss paperclips at each other…books don’t come off the shelves, but we do share opinions, we question one another, my partner/teacher that I work with….It’s good that we question one another…that we have disagreements, because then we see the other person’s point of view, as to maybe what she may see…and not as a fault in the test or in the assessment of any kind and vice versa for myself.”

In describing the benefits of assessment, teachers appreciated knowing “how much the students knew” (using tests), particularly if the student(s) had performed well.

Most teachers did not rely much on test banks to develop their test items, or the items were revised before being used because they felt that the questions in teachers’ editions or test banks were not relevant to their students’ background, reading ability, etc. Teachers reported changing tests at least once a year.

Teachers expressed need for further learning in a variety of areas of assessment. These areas ranged from creating better multiple choice exams to needing support for creating and grading portfolios. One teacher also wanted to know more about the processes used in development of standardized tests. The areas in which teachers wanted more training and knowledge were generally those in which they
reported lower self-efficacy, hence, validating their self-perceptions. In one case (Katarina), however, the quality of tests was much higher than her lower self-rating would warrant.

All teachers-taught classes comprised largely of minority students. Two teachers expressed some concern about assessing students who are English Language Learners -- their concerns were primarily about creating tests that bypassed the language barriers. One teacher was very frustrated with the way standardized tests functioned with this population of students and reported increasing her efforts to familiarize students with test-taking strategies.

All teachers made modifications to their tests for students with disabilities. These modifications typically included making tests shorter and reducing the number of options in multiple choice questions.

In addition, teachers perceived teaching and assessment as interdependent and integrated processes. They reported heavy reliance on both formal and informal assessment to inform teaching.

An emergent purpose of this study was to determine (through an evaluation of the artifacts/assessments teachers used in the past month) congruence between teacher self-efficacy in developing and grading different types of assessment and teacher use of assessment. In most cases, teachers' self-efficacy in preparation of tests seemed to be moderate to high and the reported self-efficacy was congruent with demonstrated skill in assessment. An examination of their assessment instruments and other documentation revealed some minor weaknesses like inadequate directions for taking the test, common errors in questions that resulted in clues for answering them, and not enough space to write answers.

In summary, teachers' negative beliefs about standardized assessment (self-efficacy, perceived advantages and disadvantages), their concern for their students, and their deep commitment to assessing for the sake of learning and teaching emerged from in-depth analysis of interview responses, narratives, and assessment instruments used by teachers in the month preceding the interview. The authors believe that using case studies to investigate this topic adds validity and further insight into this critical area about teaching practice. As expected with this group, teachers' knowledge about assessment is good with teachers needing additional support in developing and grading alternative assessments like portfolios and anecdotal records. Common errors in assessment artifacts were also evident, showing an imperfect (moderate) relationship between self-efficacy and demonstrated effectiveness in use of assessments. This exploratory study suggests implications for future research, particularly in the sociopolitical climate of mandated and high stakes testing, and more detailed examination of assessment procedures as used by teachers. Abrams, Pedulla, and Madaus [2003] believe that state testing (even above content standards) is the more powerful influence on teaching practices, and a great majority of teachers in state-mandated testing contexts reported that their state test has influenced them to teach in ways that oppose their own view of sound educational practice.

**IMPLICATIONS FOR TRAINING & TESTING REQUIREMENTS**

In keeping with Mertler's (2004) belief that teachers may benefit from assessment training in inservice settings, the results of this research suggest that inservice support for assessment is critical. This is particularly important because of the socio-political context of high stakes testing. It is common knowledge (among teachers) that students demonstrate their learning best in different ways. However, teachers are required to ensure that students achieve well in testing approaches that use limited formats (multiple choice and timed writing assignments). The findings from this research are that: (a) support in schools for assessment in schools is generally low, (b) the stakes are high, (c) teachers do report lower self-efficacy in a variety of areas of assessment (but mostly in multiple choice), and (d) they express a considerable frustration with standardized (akin to multiple choice) testing and the use of these scores for improving teaching and learning. The resultant cognitive dissonance and demands on teacher and student time and teacher's cognitive resources makes the need for inservice training and support imperative. This training and support may need to be individualized, where teachers can consult with an expert and ask questions about effective methods for analyzing and using data to inform instruction.

Mandated testing requirements by schools may reduce teacher flexibility and teaching effectiveness due to the time and efforts these require. Teachers have lower faith in multiple choice testing. When asked to spend more time engaged in this type of testing, they may allocate time in addition to their other assessment efforts, hence taking away from instructional time. It might be prudent to reduce this
mandated testing requirement and let teachers decide the best way to prepare students for standardized assessments and to improve student knowledge and enthusiasm about learning.

Support and training of teachers should also include demonstrations of the use of different types of assessment, particularly multiple choice tests, in making instructional decisions. Teachers who believe that such types of tests (e.g., multiple choice) play a small role in instructional decision-making, but still have to implement them in their classrooms, are likely to experience lowered morale and burnout. A related area for training could also be implementing strategies to involve parents in student assessment.

**IMPLICATIONS FOR FUTURE RESEARCH**

Future research could focus on further investigating these findings. Primarily, the relationship between teacher self-efficacy and effective use of assessment could be examined using an experimental design to determine any cause-and-effect relationships between these two variables (e.g., do programs that increase self-efficacy with specific types of assessment cause more effective use of that assessment?).

It might also be valuable to more deeply examine teachers’ beliefs about the purposes of different types of assessment. Do teachers believe that multiple choice exams do not provide much valuable information that might inform teaching and learning? Has it become an “easy-to-grade” way to assess? Do they believe that portfolios are just too time consuming and too subjective? Is it possible that teachers indeed believe that objective-type questions are to be used only to prepare students for standardized, high stakes tests?

A related area for further investigation is the amount of time teachers allocate to different assessment procedures. This could be accomplished via methods that involve observation in addition to self-report methods, because retrospective self-reports are vulnerable to many biases (including social desirability effects, memory reconstruction, and memory decay). In addition, teachers perceive teaching and assessment as integrated activities and may not be able to accurately determine the amount of time they spend on assessment alone.

Teachers expressed concern about the impact of high stakes testing on students. An area to be further explored is student and parent perceptions of high stakes assessments. Perhaps, it would be helpful to examine these perceptions in the context of lower student learning self-efficacy and the likelihood of dropping out from school. This case study research has proved valuable in setting the stage for further research in an area that is increasingly the focus of educators, parents, and politicians.

**REFERENCES**


Chase, C., Contemporary Assessment for Educators (Longman, 1999).

Gredler, M., Classroom Assessment and Learning (Longman, 1999).


