Running head: ASSESSMENT IN STATISTICS COURSES

Assessment in Statistics Courses: More Than a Tool for Evaluation

Anthony J. Onwuegbuzie

Howard University

Nancy L. Leech

University of Colorado at Denver

Correspondence should be addressed to Anthony J. Onwuegbuzie, Department of Human Development and Psychoeducational Studies, School of Education, Howard University, 2441 Fourth Street, NW, Washington, DC 20059, or E-Mail: (tonyonwuegbuzie@aol.com)

Abstract

The current assessment reform movement in statistics encourages instructors to think more broadly about cognitive measures which assess student learning. In response, statistics instructors have begun incorporating innovative methods of assessment into their courses, the most common of these procedures being authentic assessment, performance assessment, and portfolio assessment. This paper will discuss areas to consider for assessment, problems with typical assessments, and statistical authenticity for understanding student learning.

Assessment in Statistics Courses: More Than a Tool for Evaluation

As stated in the May 2000 edition of the *Educational Researcher*, the theme of the American Educational Research Association (AERA) 2001 annual meeting was "What we know and how we know it" (AERA, 2000, p. 27). Moreover, AERA called for "penetrating and weighty discussions around issues of research methodologies, rigor, standards—within every research paradigm" (AERA, 2000, p. 27). As the annual meeting theme suggests, discussions about epistemological, ontological, and axiological underpinnings of educational research are paramount. Nowhere is such dialogue as important as in the field of statistics. This importance stems from the fact that virtually every graduate student enrolled in programs representing the field of education is required to take at least one statistics and/or quantitative-based research methodology course (Mundfrom, Shaw, Thomas, Young, & Moore, 1998).

Unfortunately, for many of these students, statistics is one of the most difficult courses in their programs of study (Schacht & Stewart, 1990). Additionally, research indicates that many college students experience high levels of statistics anxiety when confronted with statistical ideas, problems, or issues, instructional situations, or evaluative situations (Feinberg & Halperin, 1978; Onwuegbuzie & Daley, 1996; Onwuegbuzie & Seaman, 1995; Roberts & Bilderback, 1980; Zeidner, 1991). The levels of statistics anxiety experienced by as many as 80% of students (Onwuegbuzie, 1998) can be so great that undertaking a statistics class is regarded by many as extremely negative, and perhaps, more importantly, as a major threat to the attainment of their degrees. In fact, as a result of anxiety, students often delay enrolling in statistics courses for as long as possible, sometimes waiting until the final semester of their degree programs--which is clearly not the optimal time to undertake such courses

(Onwuegbuzie, 1997a, 1997b; Roberts & Bilderback, 1980). Moreover, many students do not regard statistics to be a relevant or important component of their degree programs, but merely a pervasive obstacle that they must overcome in order to graduate (Gal & Ginsberg, 1994). This appears to be the case for both undergraduate and graduate students.

Students who view statistics classes as obstacle courses tend to exhibit external loci of control, coupled with overwhelming fear of failing these courses (Onwuegbuzie, DaRos, & Ryan, 1997). Indeed, using phenomenological techniques, Onwuegbuzie et al. (1997) found that failure anxiety is extremely prevalent among students enrolled in statistics classes. According to these researchers, failure anxiety comprises the following three dimensions: study-related anxiety, test anxiety, and grade anxiety. Study-related anxiety involves anxiety experienced when preparing for a test. Test anxiety pertains to anxiety experienced while taking a statistics test. Finally, grade anxiety refers to the anxiety that arises from students' expectations of their final grades. These expectations often are incongruent with reality. For some students, the expectation may be too high, whereas for others, it may be too low. In either case, it can be anxiety-inducing.

Students with one or more of these components of failure anxiety, when compared to their less-anxious counterparts, seemingly obsess with the assessment measures used by statistics instructors (Hubbard, 1997). In particular, these students tend to be preoccupied with past or upcoming in-class examinations (Onwuegbuzie et al., 1997). Consistent with this finding, using the Statistical Anxiety Rating Scale (STARS) created by Cruise and Wilkins (1980), Onwuegbuzie (1998) found students report statistically significantly higher levels of *test and class anxiety*, than the other five

dimensions of the STARS. All the effect sizes, as measured by Cohen's (1988) *d*, corresponding to these comparisons involving test and class anxiety were greater than .60.

Disturbingly, not only has statistics anxiety been found to be related negatively to statistics achievement (Elmore, Lewis, & Bay, 1993; Lalonde & Gardner, 1993; Onwuegbuzie & Seaman, 1995; Zeidner, 1991), but this construct has been reported to be the best predictor of achievement in research methodology (Onwuegbuzie, Slate, Paterson, Watson, & Schwartz, 2000) and statistics (Fitzgerald, Jurs, & Hudson, 1996) courses. Most recently, using path analytical techniques, Onwuegbuzie (2000) found that statistics anxiety, alongside achievement expectation, played a central role in the prediction of performance in statistics courses, mediating the relationship between statistics achievement and the following variables: research anxiety, study habits, course load, and the number of statistics courses taken. Moreover, using an experimental design, a causal link between statistics anxiety and course achievement has been documented (Onwuegbuzie & Seaman, 1995). Further, again using experimental techniques, students with poor examination-taking coping skills have been found to attain lower levels of performance on timed statistics examinations than do students with adequate coping skills (Onwuegbuzie & Daley, 1996).

The fact that high levels of underachievement and test anxiety prevail in statistics courses has led to calls for reform in the ways in which students are assessed in these classes (Gal & Ginsburg, 1994). Interestingly, until recently, many statistics instructors thought of assessment only in terms of testing and grading (Garfield, 1994). Indeed, because learning statistics typically was viewed as mastering a specific set of skills, formulae, vocabulary, and techniques, student assessment tended to involve in-class

tests of computational skills and rote memorization (Hawkins, Jolliffe, & Glickman, 1992). As such, items on these tests tended to examine skills in isolation of a real-life problem context and did not necessarily assess whether students fully understood statistical concepts, were able to integrate statistical knowledge to solve a novel problem, were able adequately to communicate statistical findings, or were able to communicate effectively utilizing statistical terminology (Garfield, 1994). Moreover, some students who produced a correct response to an item on these traditional statistics tests often did not understand this solution or the underlying question behind it (Jolliffe, 1991). Yet, as noted by Onwuegbuzie (2000),

> the purpose of assessment should be multifold, including the following: (1) providing information which will facilitate decisions regarding the improvement of instruction; (2) motivating and helping students to structure their learning endeavors; (3) providing individual information to students about the extent to which they are mastering the material covered; (4) reinforcing learning by providing students with indicators of what aspects of the curriculum they have not yet mastered, and on which they should focus; (5) informing instructors about how well the classes appear to understand particular topics and what topics should be re-introduced; (6) providing diagnostic information to instructors about individual students' strengths and weaknesses in understanding new material; and (7) providing an overall indicator of students' performance levels (Busk, 1998; Garfield, 1994; National Council of Teachers of Mathematics [NCTM], 1993; Webb & Romberg, 1992). (p. 322)

It is unlikely that traditional in-class assessments can meet all of these goals. For example, when students receive a total score for their responses to an in-class test, this

summary statistic is unable to inform students as to what aspects of the curriculum they have not yet mastered, nor, in the absence of a thorough item analysis, does such a statistic inform the instructor of students' areas of weakness. Moreover, as the goals and objectives for the teaching of educational statistics continue to evolve as we progress through the 21st century, traditional assessments are more apt to be misaligned to desired student outcomes.

Rather, as envisioned and advocated by the National Council of Teachers of Mathematics (NCTM), measures of statistics performance should be an active process that yields information about students' progress towards the achievement of course goals and objectives on an on-going process. According to NCTM (1993), when the information derived from assessment instruments is consistent with course goals and is used effectively to inform instruction, it serves to promote student learning as well as to monitor it. In fact, assessments should be used not only to provide information to students and instructors alike, but also in research on teaching and learning statistics, as well as in assessing the efficacy of different curricula or pedagogical techniques (Garfield, 1998).

In light of the aforementioned criteria, a comprehensive approach to assessment is needed, beyond that of traditional testing and grading (Onwuegbuzie, 2000). Encouragingly, rather than being an activity distinct from instruction, as until recently has been the case in statistics courses, assessment is now being utilized as an integral part of both teaching and learning (Mathematical Sciences Education Board, 1993). Thus, the current assessment reform movement in statistics encourages instructors to incorporate cognitive measures that assess student learning more extensively (Garfield, 1994; Lesh & Lamon, 1992; Romberg, 1992). In response, statistics instructors have

begun utilizing creative methods of assessment in their courses (Onwuegbuzie, 2000).

Before deciding on the method(s) of assessment to use in a statistics class, the instructor must reflect upon a myriad of considerations. These considerations comprise the context in which the course is taught, the desired content of the course, and the preferred pedagogical style of the instructor. The relationships among these variables are presented in Figure 1.

Insert Figure 1 about here

Indeed, as can be seen from this figure, the context of teaching statistics represents the first consideration for statistics instructors. That is, before deciding how to assess statistics learning, the instructor should take into consideration the context in which the class is taught. Next, the educator should then simultaneously take into account the intended content of the course (i.e., curriculum) and her/his pedagogical style. After considering these three components, the instructor is now ready to design the course assessments. However, it should be noted that the relationship among the content, pedagogical style, and assessment is somewhat recursive. That is, just as the content and pedagogical style influence the eventual assessment tools used in the statistics course, the type of assessment techniques incorporated can influence both the content and pedagogical style. Considerations regarding the context, content, and pedagogical style are discussed below.

Considerations in the Area of Assessment

Decisions about statistics assessments cannot be made without a complete consideration of the context, content, and pedagogical style. Once these considerations

have been made, the statistics instructor is then ready to design an assessment package. There are five basic considerations and three dimensions necessary to consider when thinking about assessments for statistics. The fundamental decisions, as noted by Garfield (1994), include the following five dimensions: (a) what to assess, (b) the purpose of assessment, (c) how to assess it, (d) who will undertake the assessment, and (e) the action to be taken by the instructor and the nature of feedback given. Clearly, these five facets are dependent on one another. The first component, what to assess, comprises concepts, skills, applications, attitudes, and beliefs (Garfield, 1994). The second consideration, the purpose of assessment, forces the instructor to reflect upon his/her philosophical underpinnings for assessing statistics learning. The third consideration, namely, how to assess statistics learning, depends largely on the purpose of the assessment. For example, if the purpose of the assessment is to evaluate students' ability to communicate statistical findings to groups of individuals, then the instructor is more likely to require oral presentations.

The fourth consideration of the assessment framework is who will undertake the assignment. Possible administrators are the course instructor, peers, and the students themselves. Although the former prevails, it is important for students to learn how to evaluate and to apply their own knowledge and skills (Garfield, 1994). One way of helping students to engage in self-assessment is via scoring rubrics (Wilson & Onwuegbuzie, 1999). These rubrics allow students to apply scoring criteria to their own work, as well as to their peers, so that they can learn how their ratings compare to those of their instructor (Wilson & Onwuegbuzie, 1999). Other ways of assisting students to self-assess their work is by providing them with model papers and exemplars of good performance in advance. Such models allow students to know the performance

standards expected by the statistics teacher (Garfield, 1994).

The fifth and final consideration is the action that the instructor intends to take based on the results of the assessment and the nature of feedback provided to students. According to Garfield (1994, paragraph 26), “this is a crucial component of the assessment process that provides the link between assessment and improved student learning.” These five considerations then form a useful framework for designing assessment tools in statistics courses.

Assessments in statistics courses can be conceptualized along the following three dimensions: *time*, *memorization*, and *response delivery*. With respect to time, examinations lie on a continuum from “timed” to “untimed.” The extreme end of the timed continuum includes speeded statistics examinations, whereas the extreme end of the “untimed” continuum include in-class examinations with no time limits. With regard to memorization, statistics assessments range on a continuum from examinations with no supporting material allowed to examinations with unlimited supporting material allowed (e.g., take-home examinations). Finally, with respect to response delivery, statistics examinations lie on a continuum from written to oral. These three dimensions, time, memorization, and response delivery, are useful in characterizing different types of assessments in statistics courses. Interestingly, Onwuegbuzie (2000), who investigated the methods of statistics assessments that students most prefer via an exploratory factor analysis, found that students’ preferred assessment styles centered around the following three themes: oral presentations, supporting material, and timed/no support/authentic assessments. After taking into account the five considerations of what to assess, the purpose of assessment, how to assess it, who will undertake the assessment, and the action to be taken by the instructor and the nature of feedback

given, and the three dimensions of time, memorization, and response delivery, an appropriate assessment can be chosen.

Problems With Typical Assessments

There are many assessments that have typically been used in statistics courses. These include quizzes, in-class examinations, take-home examinations, term projects with peer reviews, portfolios, simulations, oral presentations, computer laboratory components, minute papers (i.e., short description compiled by students as to what they have learned and not understood during class), attitude surveys, journal entries, performance assessments, and authentic assessments (Cobb, 1993; Garfield, 1994; Onwuegbuzie, 2000). There are problems with each of these typical assessments, but in-class examinations tend to present the most problems for students in statistics courses.

In-class examinations, which are the most traditional types of assessment, can involve multiple-choice, computational, short-answer, or essay items. These examinations tend to involve time limits. However, such assessment formats have been found to be problematic in statistics courses. In particular, Onwuegbuzie and Seaman (1995) found that graduate students with high levels of statistics test anxiety who were randomly assigned to a statistics examination that was administered under timed conditions tended to have lower levels of performance than did their high-anxious counterparts who were administered the same test under untimed conditions.

In a follow-up experimental investigation among female college students, Onwuegbuzie (1995) reported a statistically significant interaction between statistics test anxiety and type of examination (i.e., timed vs. untimed), with high-anxious female students showing a greater decrement in performance than did low-anxious female

students in the untimed examination condition. Onwuegbuzie interpreted these results within conceptual frameworks developed by Hill (1984) and Wine (1980), who suggested that differences between high- and low-anxious students in evaluative situations are due to differences in motivational dispositions and attentional foci, respectively. Thus, instructors should be cognizant that in administering timed examinations, they are not only measuring statistics ability but also levels of anxiety. Rather, instructors should consider administering untimed examinations. Interestingly, Onwuegbuzie (2000) found that examinations that are untimed and in which supporting material is allowed are regarded by the majority of students as inducing the least amount of anxiety, as increasing levels of performance, and as promoting higher-order thinking.

Statistical Authenticity

The problems with traditional methods of assessment leave statistics instructors with the dilemma of not knowing how to assess accurately students' level of understanding. Derry, Levin, and Schuable (1995) identified statistical authenticity as one possible method of assessment. An issue with most statistics courses and assessments is that they are detached from real-life problem solving (Derry et al., 1995). Thus, as advocated by Derry et al. (1995), instructors should focus on what they term "statistical authenticity" (p. 53). Focusing on statistical authenticity in assessment can create a better method for understanding students' level of understanding.

Statistical authenticity has been defined by Derry et al. (1995) as lying along the following two dimensions: cultural relevance and social activity. The relevance component refers to the extent to which statistical reasoning is linked to meaningful, everyday real-life problems that are deemed to be important by society. The social

activity dimension pertains to “the extent to which learning emerges from active conceptualizing, negotiation, and argumentation” (p. 53). Accordingly, Derry et al. call for instructors to design courses that reach the highest possible level of statistical authenticity. Apparently, the point at which the highest level of statistical authenticity is reached is when both cultural relevance and social activity are maximized. This occurs through

> (a) instruction that uses examples, illustrations, and demonstrations that are relevant to the cultures to which students belong or hope to belong (cultural relevance); and (b) mentored participation in a social, collaborative problem-solving context, with the aid of such vehicles as group discussion, debate, role-playing, and guided discovery (social activity). (p. 54).

Perhaps the best way of attaining statistical authenticity is by administering performance assessments and authentic assessments. Performance assessment involves providing students with tasks, projects, or investigations, then formally evaluating the products that emerge in order to determine what students have learned and how they can apply this knowledge (Stenmark, 1991). Accordingly, performance assessment tasks should reflect important, meaningful, interesting, and thought-provoking performances that are linked to desired real-life student outcomes (Fuchs, 1995; Wiggins, 1989; Worthen, 1993). Moreover, performance assessment involves blending content with process and major concepts with specific problems (Baron, 1990). As such, performance assessments assess what students can do, as well as what they know (Hutchinson, 1995), having observing, documenting, and analyzing student work at its core (Davey & Neill, 1991).

As noted by Hutchinson (1995), performance assessments can utilize flexible

time frames, open-ended formats, and cooperative and collaborative learning techniques. Also, modifications can be made to performance assessments that are based on students' abilities, experiences, and skills (Lam, 1995). According to Elliot (1995), when performance assessments are used, students' level of performance can be improved by the following: (a) selecting assessment tasks that are aligned clearly and are connected to the material being taught; (b) specifying clearly the scoring criteria for the assessment task to students prior to attempting the task; (c) providing students with explicit statements of standards and/or various models of acceptable performance before they attempt a task; (d) encouraging students to undertake self-assessments of their performances; and (e) interpreting students' performances by comparing them to those of other students, as well as to standards which are developmentally appropriate.

Authentic assessments represent a method of collecting information regarding students' understanding in contexts that reflect real-life, everyday situations, and which challenge students to apply what they have learned in their courses in authentic settings (Archbald & Newmann, 1988). This method of assessment provides students with expectations about what will be assessed, as well as standards to be met in realistic contexts. Also, they present students with information about where they are in relation to where they need to be (Lankard, 1996). More specifically, according to Wiggins (1990), "authentic assessments present the student with the full array of tasks that mirror the priorities and challenges found in the best instructional activities: conducting research; writing, revising and discussing papers; providing an engaging oral analysis of a recent political event; collaborating with others on a debate, etc." (p. 2). As such, authentic assessments help students to be effective performers with acquired knowledge (Wiggins, 1990).

As contended by Onwuegbuzie (2000, p. 323), “both authentic assessments and performance assessments provide a basis for statistics instructors to evaluate both the effectiveness of the process (i.e., the procedure used) and the product resulting from the performance of a task (e.g., a completed report).” Whereas in-class examinations typically measure factual knowledge, in performance and authentic assessments, there is often no single correct or even best solution. Rather, there may be several viable performances and solutions. The use of performance assessments and authentic assessments is consistent with the report of the ASA Section on Statistical Education Committee on Training of Statisticians for Industry (1980), which strongly encourages the development of students’ practical skills. Indeed, the authors of this document note that many programs in the United States fail to achieve this goal.

It should be noted that although authentic assessments and performance assessments are similar in their measurement of process and product, as noted by Linn and Gronlund (1999), they are distinct. Indeed, the essential difference between authentic assessments and performance assessments is that the former must involve the application of knowledge in authentic settings, whereas performance assessments do not emphasize the practical application of the tasks in real-world settings. Simply put, in authentic assessment, authenticity is required, whereas in performance assessment, authenticity is usually only approximated (Linn & Gronlund, 1999).

For example, in a statistics course, performance assessment could involve analyzing and interpreting fake data or data provided by the instructor. At Valdosta State University (Georgia) and at Howard University (Washington, DC), doctoral students enrolled in statistics courses are required to complete a statistics notebook, whereby students are asked to analyze real data provided by the instructor for every

statistical technique taught in the course. Students also are required to write up formally the results in the same manner as would appear in a published journal article. Detailed scoring rubrics are provided for this performance-based assessment (Wilson & Onwuegbuzie, 1999).

On the other hand, authentic assessment could involve students collecting, analyzing, and interpreting real data, and then submitting it to a journal for possible publication, as is required in some doctoral programs. At Valdosta State University, doctoral students also are required to complete a mini-dissertation using real data. The goal of the mini-dissertation is to allow students to practice formulating research questions and hypotheses, conducting reviews of the literature, and collecting, analyzing, and interpreting quantitative data. Mini-dissertations must contain all the major elements of the five chapters of a traditional dissertation. It is expected that, upon completion of the mini-dissertation, students will be familiar with every aspect of the dissertation process. As such, the mini-dissertation is intended to play a major role in *demystifying the dissertation process*. Detailed scoring rubrics were used to help students understand the process (Wilson & Onwuegbuzie, 1999).

Additionally, at this same institution, students are required to conduct a 15-minute professional presentation of her/his mini-dissertation. Students must be dressed in a professional manner in order to simulate real professional conferences. The goal of these oral presentations is to give students an opportunity to present their research findings in a formal setting. Again, detailed feedback is provided utilizing a scoring rubric.

Many doctoral students at Valdosta State University turn their mini-dissertations into journal-ready articles the following semester. In fact, encouragingly, within the last

several months alone, more than one dozen doctoral candidates at Valdosta State University have had their mini-dissertations published in reputable nationally refereed journals. Additionally, more than 20 students have presented findings from their mini-dissertations at professional conferences. Similarly, at Howard University, several students have published articles in nationally refereed journals based on empirical research papers written in statistics and measurement courses.

Interestingly, Onwuegbuzie (2000) found that students tended to rate performance assessments more highly than they did other examination formats. Further, authentic assessments were rated as best promoting higher-order thinking. Also, both performance and authentic assessments promote active learning, which, in turn, promotes students' sense of responsibility (Cobb, 1993). Mini-dissertations also guarantee that the statistics instructor also will experience learning in the course as students analyze real data that have yet to be analyzed. Thus, performance assessments and authentic assessments present viable alternatives to traditional assessments.

Whatever assessment tools are used, it is imperative that they receive as prompt feedback as possible (Cross, 1987; McKeachie, Pintrich, Lin, & Smith, 1986). Indeed, prompt feedback has been found to be related to student achievement and satisfaction (Dunkin, 1986; McKeachie et al., 1986). Also, it is advisable to use several assessment tools simultaneously in order to provide data about teaching and learning that are triangulated.

Summary and Conclusions

Research suggests that the statistical preparation skills among non-quantitative majors are perceived by many statistics instructors to be inadequate (Curtis & Harwell,

1996). Moreover, as noted by Lomax and Moosavi (1998), “the pedagogy in the discipline of statistics has not changed nearly as quickly as most disciplines. Doctoral students in statistics are then mentored by these same professors who, in turn, engage in these same teaching practices. Thus, the cycle seems to perpetuate itself” (p. 3). This paper has attempted to help break this cycle by providing information to statistics instructors and others about assessments in statistics courses.

Students prepare for examinations in ways that reflect how they believe they will be tested (Crooks, 1988; McKeachie, 1986; Wergin, 1988). For example, if they expect an examination that focuses on facts, they will memorize specifics. Conversely, if they expect an examination that necessitates problem solving or integrating knowledge, students will strive to understand and to apply information (Busk, 1998).

As noted by Garfield (1998), items on traditional statistical examinations often lack adequate context and tend to focus on the regurgitation and application of statistical formulae and the accuracy of statistical computations. Such items typically are scored dichotomously and, as such, do not sufficiently mirror the nature of students’ reasoning and problem solving, thereby, at best, providing limited information about students’ level of substantive statistical understanding (Gal & Garfield, 1997; Garfield, 1998).

Further, if statistics examinations emphasize rote learning, students can only provide a correct answer if the question is posed in exactly the form in which they have learned it. As such, these students will not be able to transfer their knowledge to solve novel problems that occur outside the textbook and outside the classroom. Additionally, students tend to forget techniques that they learned without understanding (Hubbard, 1997). Thus, tests should be designed that attempt to measure what students

understand and not what they can calculate. Items on such tests should have the right amount of ambiguity so that students are forced to reflect on a range of competing responses. Most importantly, students should not be led to believe that statistics examinations involve tricks designed to confuse students, and that they necessitate performing statistical gymnastics.

Rather than viewing assessment as a tool that drives instruction and learning as do many educators (e.g., Cobb, 1993; Garfield, 1994), we should treat assessment as an essential component of statistics classes that influences and is influenced by the context, content, and pedagogical style. In other words, context, content, pedagogical style, and assessment should be viewed as representing an interactive, iterative, and recursive process in every statistics class.

As we continue our passage through the 21st century, it is important that all statistics instructors examine closely the assessment techniques currently utilized in their classrooms. In reflecting on assessment methods, it is imperative that statistics instructors operate under the assumption that “however motivating we make the instruction, some students will fail to be motivated to take a real interest in our discipline unless we also make changes to our methods” (Hubbard, 1997, paragraph 2). In so doing, it is important that statistics educators view assessments not only as tools for assigning grades for performance, but also as instruments for providing information on how to improve teaching and learning (Busk, 1998; Garfield, 1994; NCTM, 1993; Webb & Romberg, 1992). In making any adjustments to the assessment regime, it is important not to do so all at once, but to use a more piecemeal approach (Garfield, 1994). Further, statistics instructors should not attempt to make wholesale assessment changes in isolation of other colleagues. In fact, instructors should seek advice and

feedback from trusted colleagues. Instructors also should confer with fellow statistics teachers at other institutions to keep abreast with the latest assessment techniques. Resources such as the *Journal of Statistics Education* could be consulted for ideas of different assessment approaches.

Once implemented, statistics instructors should be clear and explicit to their students about how and why they are being assessed (Chance, 1997; Stenmark, 1991). In evaluating the effectiveness of assessment methods, instructors should assess students' beliefs and attitudes towards these measures (Gal & Ginsberg, 1994; Onwuegbuzie, 2000). Further, once the effectiveness of an assessment tool has been documented, this information should be disseminated to as many other statistics instructors as possible via published articles, paper presentations, symposia, and the like.

By carefully designing assessment techniques, statistics instructors will help to create what Derry et al. (2000) phrase as a "culture of expert practice in the classroom" (p. 750) in which usable knowledge is created, and whereby students learn spontaneously to activate appropriate statistical ideas and procedures. Finally, in utilizing innovative assessment techniques, statistics instructors should be aware of the fact that reform-based courses are difficult to implement as a result of social cognition theory, which predicts that cognition and motivation are strongly affected by the broader social and institutional context in which they occur (Derry et al., 2000; Nicolopoulou & Weintraub, 1998). Unfortunately, traditional courses and techniques are maintained by a long history of being positioned within potent cultural contexts that often tend to counteract reform-based instruction. Thus, new assessment techniques initially may be difficult to gain widespread approval among students and faculty because they are more

time-consuming for instructors, more challenging for students, and are inconsistent with the goals and expectations held by many individuals. As noted by Cobb (1993), resistances to innovation typically are merely symptoms of other problems often associated with change. Cobb (1993) contends that by using the Total Quality Management framework (see Walton, 1986), these problems can be addressed effectively by attending to logistics, by being explicit about goals, objectives, and standards, and by collecting evaluation data from students. Regardless, by initiating and maintaining discourse about innovative statistical techniques, statistics instructors will be taking an important step in promoting reform-based assessment initiatives in statistics courses.

References

American Educational Research Association. (2000). American Educational Research Association 2001 Annual Meeting Call for Proposals. *Educational Researcher, 29(*4), 27-41.

Archbald, D., & Newmann, F. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school*. Reston, VA: National Association of Secondary School Principals.

ASA Section on Statistical Education Committee on Training of Statisticians for Industry (1980). Preparing statisticians for careers in Industry. *The American Statistician, 34*, 65-75.

Baron, J. B. (1990, April). *How science is tested and taught in elementary school science classrooms: A study of classroom observations and interviews*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Busk, P. (1998, April). *Using assessment techniques for facilitating student learning in statistics course*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Chance, B. L. (1997). Experiences with authentic assessment techniques in an introductory statistics course. *Journal of Statistics Education* [On-line serial], *5*(3). Available E-mail: archive@jse.stat.ncsu.edu. Message: send jse/v5n3/chance.

Cobb, G. (1993). Reconsidering statistics education: A National Science Foundation Conference. *Journal of Statistics Education* [On-line serial], *1*(1). Available E-mail: archive@jse.stat.ncsu.edu. Message: send jse/v1n1/cobb.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.).

Hillsdale, NJ: Erlbaum.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58,* 438-481.

Cross, K.P. (1987). Teaching for learning. *AAHE Bulletin, 39*(8), 3-7.

Cruise, R. J., & Wilkins, E. M. (1980). *STARS: Statistical Anxiety Rating Scale*. Unpublished manuscript, Andrews University, Berrien Springs, MI.

Curtis, D. A., & Harwell, M. (1996, April). *Training graduate students in educational statistics: A national survey*. Paper presented at the annual meeting of the American Educational Research Association, New York: New York.

Davey, L., & Neill, M. (1991). *The case against a national test. ERIC/TM Digest*. (ERIC Document Reproduction Service No. ED 338 703)

Derry, S., Levin, J. R., & Schuable, L. (1995). Stimulating statistical thinking through situated simulations. *Teaching of Psychology, 22,* 51-57.

Dunkin, M. J. (1986). Research on teaching in higher education. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp. 754-777). New York: MacMillan.

Elliott, S. N. (1995). *Creating meaningful performance assessments. ERIC Digest E531*. (ERIC Document Reproduction Service No. ED 381 985)

Elmore, P. B., Lewis, E. L., & Bay, M. L. G. (1993, April). *Statistics achievement: A function of attitudes and related experience.* Paper presented at the annual meeting of the American Educational Research Association. (ERIC Document Reproduction Service No. 360 324)

Feinberg, L., & Halperin, S. (1978). Affective and cognitive correlates of course performance in introductory statistics. *Journal of Experimental Education, 46*(4), 11-18.

Fitzgerald, S. M., Jurs, S., & Hudson, L. M. (1996). A model predicting statistics achievement among graduate students. *College Student Journal, 30,* 361-366.

Fuchs, L. S. (1995). *Connecting performance assessment to instruction: A comparison of behavioral assessment, mastery learning, curriculum-based measurement, and performance assessment. ERIC Digest E530*. (ERIC Document Reproduction Service No. ED 381 984)

Gal, I., & Garfield, J. (Eds.) (1997). *The assessment challenge in statistics education*. Amsterdam: IOS Press.

Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education [On-line serial], 2(*2). Available E-mail: archive@jse.stat.ncsu.edu. Message: send jse/v2n2/gal.

Garfield, J. B. (1994). Beyond testing and grading: Using an assessment to improve student learning. *Journal of Statistics Education* [On-line serial], *2*(1). Available E-mail: archive@jse.stat.ncsu.edu. Message: send jse/v2n1/garfield.

Garfield, J. B. (1998, April). *Challenges in assessing statistical reasoning*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Hawkins, A., Jolliffe, F., & Glickman, L. (1992). *Teaching statistical concepts*. Harlow, Essex, England: Longman Group UK Limited.

Hill, K. T. (1984). Debilitating motivation and testing: A major educational program, possible solutions, and policy applications. In R.E. Ames and C. Ames (Eds.), *Research on motivation in education*, *1* (pp. 245-274.). New York: Academic

Press.

Hubbard, R. (1997). Assessment and the process of learning statistics. *Journal of Statistics Education* [On-line serial], *2*(2). Available E-mail: archive@jse.stat.ncsu.edu. Message: send jse/v5n1/hubbard.

Hutchinson, N. L. (1995). *Performance assessments of career development: ERIC Digest*. (ERIC Document Reproduction Service No. ED 414 518)

Jolliffe, F. (1991). Assessment of the understanding of statistical concepts. In D. Vere-Jones (Ed.), *Proceedings of the third international conference on teaching Statistics* (pp. 461-466). Otago, NZ: Otago University Press.

Lalonde, R. N., & Gardner, R. C. (1993). Statistics as a second language? A model for predicting performance in psychology students. *Canadian Journal of Behavioral Science, 25,* 108-125.

Lam, T. C. M. (1995). *Fairness in performance assessment. ERIC Digest*. (ERIC Document Reproduction Service No. ED 391 982)

Lankard, B. A. (1996). *Acquiring self-knowledge for career development. ERIC Digest No. 175*. (ERIC Document Reproduction Service No. ED 399 414)

Lesh, R., & Lamon, S. (1992). *Assessment of authentic performance in school mathematics*. Washington, DC: AAAS.

Linn, R. L., & Gronlund, E. (1999). *Measurement and assessment in teaching*. Englewood Cliffs, N.J.: Prentice Hall.

Lomax, R. G., & Moosavi S. A. (1998, April). *Using humor to teach statistics: Must they be orthogonal?* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Mathematical Sciences Education Board (1993). *Measuring what counts: A conceptual guide assessment*. Washington, DC: National Academy Press.

McKeachie, W. J. (1986). *Teaching tips* (8th ed.). Lexington, MA: Heath.

McKeachie, W. J., Pintrich, P. R., Lin, Y., & Smith, D. (1986). *Teaching and learning in the college classroom: A review of the research literature.* Ann Arbor, MI: National Center to Improve Postsecondary teaching and Learning, University of Michigan.

Mundfrom, D. J., Shaw, D. G., Thomas, A., Young, S., & Moore, A. D. (1998, April). *Introductory graduate research courses: An examination of the knowledge base*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

National Council of Teachers of Mathematics (1993). *Assessment standards for school mathematics: Working draft*. Reston, VA: NCTM.

Nicolopoulou, A., & Weintraub, J. (1998). Individual and collective representations in social context: A modest contribution to resuming the interrupted project of a sociocultural development psychology. *Human Development, 41*, 215-235.

Onwuegbuzie, A. J. (1995). Statistics test anxiety and women students. *Psychology of Women Quarterly, 19*, 413-418.

Onwuegbuzie, A. J. (1997a). Writing a research proposal: The role of library anxiety, statistics anxiety, and composition anxiety. *Library & Information Science Research 19,* 5-33.

Onwuegbuzie, A. J. (1997b). The teacher as researcher: The relationship between enrollment time and achievement in a research methodology course. *Reflection*

and Research, 3(1)[On-Line].Available: http://www.soe.gonzaga.edu/rr/v3n1/tony.html

Onwuegbuzie, A. J. (1998). The dimensions of statistics anxiety: A comparison of prevalence rates among mid-southern university students. *Louisiana Educational Research Journal, 23*, 23-40.

Onwuegbuzie, A. J. (2000). Attitudes toward statistics assessments. *Assessment & Evaluation in Higher Education, 25, 321-339.*

Onwuegbuzie, A. J., & Daley, C. E. (1996). The relative contributions of examination-taking coping strategies and study coping strategies on test anxiety: A concurrent analysis. *Cognitive Therapy and Research, 20*, 287-303.

Onwuegbuzie, A. J., DaRos, D., & Ryan, J. (1997). The components of statistics anxiety: A phenomenological study. *Focus on Learning Problems in Mathematics, 19*, 11-35.

Onwuegbuzie, A. J., & Seaman, M. (1995). The effect of time and anxiety on statistics achievement. *Journal of Experimental Psychology, 63,* 115-124.

Onwuegbuzie, A. J., Slate, J., Paterson, F., Watson, M., & Schwartz, R. (2000). Factors associated with underachievement in educational research courses. *Research in the Schools, 7*, 53-65.

Roberts, D. M., & Bilderback, E. W. (1980). Reliability and validity of a statistics attitude survey. *Educational and Psychological Measurement, 40*, 235-238.

Romberg, T. (Ed.) (1992). *Mathematics assessment and evaluation: Imperatives for mathematics education*. Albany, New York: State University of New York Press.

Schacht, S., & Stewart, B. J. (1990). What's funny about statistics? A technique for

reducing student anxiety. *Teaching Sociology, 18*, 52-56.

Stenmark, J. (1991). *Mathematics assessment: Myths, models, good questions, and practical suggestions*. Reston, VA: NCTM.

Walton, M. (1986). *The Deming Management Method*. New York: The Putnam Publishing Group.

Webb, N., & Romberg, T. (1992). Implications of the NCTM standards for mathematics assessment. In T. Romberg (Ed.), *Mathematics assessment and evaluation: Imperatives for mathematics education* (pp. 37-60). Albany, NY: State University Mathematics Education.

Wergin, J. E. (1988). Basic issues and principles in classroom assessment. In J. H. McMillan (Ed.), *Assessing students' learning*. New Directions for Teaching and Learning, no. 34 (pp. 5-17). San Francisco, CA: Jossey-Bass.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 70*, 703-713.

Wiggins, G. (1990). *The case for authentic assessment. ERIC Digest*. (ERIC Document Reproduction Service No. ED 328 611)

Wilson, V., & Onwuegbuzie, A.J. (1999, November). *Improving achievement and student satisfaction through criteria-based evaluation: Checklists and rubrics in educational research courses.* Paper presented at the annual conference of the Mid-South Educational Research Association, Point Clear, AL.

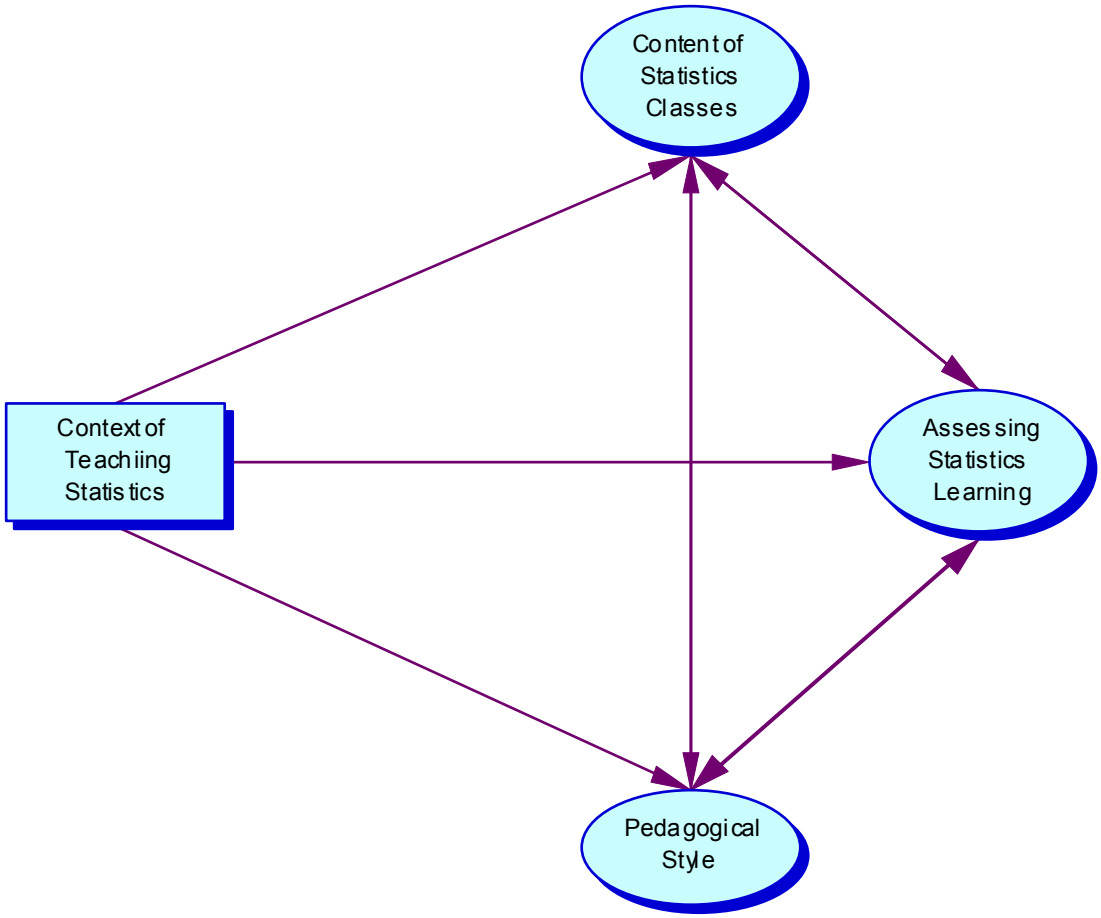
Wine, J. (1980). Cognitive-attentional theory of test anxiety. In I. G. Sarason (Ed.), *Test anxiety: Theory, research and applications* (pp. 349-385). Hillsdale, NJ: Erlbaum.

Worthen, B. R. (1993). Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan, 74*, 444-48.

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students – some interesting parallels. *British Journal of Educational Psychology, 61*, 319-328.

Figure Caption

Figure 1. Model of teaching and learning statistics.


Content of
Statistics
Classes
Context of
Teachiing
Statistics
Assessing
Statistics
Learning
Pedagogical
Style