Assessment as a Tool for Learning Statistics

Anthony J. Onwuegbuzie
Valdosta State University

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. Thus, this paper will provide a typology of different effective ways of assessing performance in statistics classes for the various contexts (e.g., undergraduate vs. master’s vs. doctoral), content (e.g., measurement vs. evaluation vs. research design), and pedagogical styles (e.g., web-based vs. traditional; theory vs. concept vs. application).
Assessment as a Tool for Learning Statistics

As stated in the May 2000 edition of the *Educational Researcher*, the theme of the American Educational Research Association (AERA) 2001 annual meeting is “What we know and how we know it” (AERA, 2000, p. 27). Moreover, AERA calls 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, 1998a) 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;
Robert & 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, typically exhibiting an overwhelming fear of failing these courses (Onwuegbuzie, DaRos, & Ryan, 1997). Indeed, using phenomenological techniques, Onwuegbuzie et al. (1997) found that failure anxiety was extremely prevalent among students enrolled in statistics classes. According to these researchers, failure anxiety comprises the following three dimensions: study-related anxiety, test anxiety, 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’ expectation of their final grade. 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 seemingly obsess with the assessment measures used by statistics instructors compared to their less-anxious counterparts (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; Cruise & Wilkins, 1980), Onwuegbuzie (1998a) found students to report significantly higher levels of test and class anxiety, than the other five dimensions of the STARS. All the effect sizes 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 (2000a) 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, a causal link between statistics anxiety and course achievement has been documented (Onwuegbuzie & Seaman, 1995). Further, students with poor examination-taking coping skills have been found to attain lower levels of performance on 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 (2000a, p. 322), 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).

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 enter 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, 2000a). 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, 2000a).
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. 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.
context in which this teaching occurs” (p. 1). As can be seen in Figure 2, issues related to the context of teaching statistics contains many facets. First and foremost, the statistics instructor should consider the type of institution in which the course is being taught (Figure 2). For example, a Research University likely is significantly different than is a traditional Teaching University with respect to the level of student (e.g., Master’s-vs. doctoral-level), type of student body (e.g., statistics vs. non-statistics major), diversity of student body, number of statistics courses in students’ program of study, levels of statistics courses (e.g., introductory vs. advanced), status of statistics course (e.g., required vs. elective), and the competence and experience of the statistics instructor. Each of these components, in turn, help to determine the goal of the statistics course.

As part of determining the goal of a statistics class, the statistics teacher must decide whether his/her students should be prepared to be consumers or producers of educational research. For instance, if the instructor decides to help students become consumers of research, then he/she may be more likely to focus on the theoretical or conceptual aspects of statistics. Conversely, to prepare students to be producers of educational research, the statistics teacher likely may devote at least some of the course to the instruction of statistical applications, including the use of computer (statistical) software. Thus, all of these facets play an important role in determining the content and pedagogical style, as well as the assessment tools used.

Wilson (2001) noted that “no discussion of the context of teaching statistics
would be complete without acknowledgement of the anxiety that students bring to class” (p. 2). Indeed, because the majority of students enrolled in statistics classes typically experience high levels of statistics anxiety (Onwuegbuzie, in press-a), and that anxiety can debilitate statistics achievement (Onwuegbuzie & Seaman, 1995), instructors should be cognizant of how the statistical milieu might affect students’ levels of anxiety.

In attempt to reduce levels of statistics anxiety in classes, several researchers have advocated the use of humor in statistics classes (Lomax & Moosavi, 1998; Schacht & Stewart, 1990; Wilson, 1999). Interestingly, in a study of education and business students in graduate statistics courses, Wilson (1996) found that students deemed humor and testing procedures (such as open book/open note testing) to be somewhat effective in reducing their anxiety levels. Moreover, Schacht and Stewart (1990) advocated teaching gimmicks in statistics classes. These gimmicks include using students as the source from which data are collected and allowing students to create the statistical application. Sgoutas-Emch and Johnson (1998) found journal writing to be effective in reducing levels of anxiety, although these authors did not find a statistically significant decrease in anxiety levels.

Most recently, utilizing Onwuegbuzie’s (1998b) finding of a relationship between hope and statistics anxiety, Dilevko (2000) advocated that statistics and research methodology class activities attempt to assist students in understanding the course objectives, as well as being aware of the goal of statistics in order to control their own learning objectives. Dilevko contended that statistics anxiety can be reduced by improving students’ perceived worth of statistics and by decreasing their fear of applying statistical knowledge and principals. According to Dilevko, a two-pronged approach should be employed to reduce statistics anxiety, namely: (a) using current news stories and similar
sources to introduce and to explain basic statistical concepts and methodological issues in research; and (b) targeting their fear of application of statistics concepts by introducing students to older articles about subjects of interest, and asking them to read, to understand, and to critique these articles, and to suggest how the research projects described in them could be modified, expanded, and updated. Dilevko asserted that by utilizing these two strategies, the importance of statistics in everyday life can be demonstrated through class discussion of interesting events reported in the popular press. Unfortunately, although appealing, Dilevko did not provide any research evidence of the efficacy of his strategies.

The Content of Statistics Courses

Once the context of the statistics course has been identified, the statistics instructor has to determine the content of the class. Here the context is essential for planning the statistics curriculum. For example, the content likely will be different for undergraduates versus graduates, as well as for master's versus doctoral students. Similarly, the content is likely to be more advanced for statistics- and research-based majors than for their counterparts. Likewise, different curricula are needed for introductory classes than for intermediate or advanced statistics classes. Whatever the context, statistics instructors should be cognizant that they “do not have enough contact hours with [their] students to cover everything that they need to know to conduct high quality research and to understand the latest developments in the field” (Johnson, 2001, p. 2). Thus, as noted by Johnson (2001), statistics educators “must make many hard decisions about what to include and exclude in [their] courses” (p. 2).

The fact that higher standards in research and statistics (e.g., Wilkinson & the Task Force on Statistical Inference, 1999; The American Statistical Association, 1999) currently prevail, coupled with the fact that our knowledge base in the areas of
quantitative-based methodology has rapidly expanded in recent year, necessitate that students, particularly at the doctoral level, take more research methods and statistics courses. Yet, it appears that doctoral curricula and the like have less and less room for research design, statistics, and measurement (Thompson, 1998a). In fact, at some institutions, such as at Valdosta State University, students are required to take only one statistics course. At such institutions, choice of topics to be taught is extremely critical (Elmore & Woehlke, 1998).

Several additional factors play a role in the content that emerges, including the following: administrator support, time, resources, interest in statistics/research, willingness to change, willingness to keep up with the statistics literature, knowledge of the best statistical practices, confidence in statistics/research, experience of teaching statistics, level of instructor anxiety, and competence in teaching statistics. Figure 3 provides a conceptual model of how these factors are related.

Especially important components of this model are awareness of the most current statistical theory and applications and a willingness to change (Johnson, 2001). As noted by Johnson (2001, p. 20), “it is essential that professors of educational research and statistics keep up on the latest developments in methodology and statistics so that [they] can pass this information along to [their] graduate students ‘in training’” [emphasis in original]. Indeed, a statistics instructor who is not well versed with the best statistical practices and/or is unwilling to disseminate these practices is unlikely to develop a cutting-edge curriculum. Unfortunately, the teaching of inappropriate or
Assessment As a Tool for Learning Statistics

out-dated statistical techniques can lead to students being developed who conduct seriously-flawed theses, dissertations, and other types of research studies. Disturbingly, statistical training has changed little during the last two decades (Aiken, West, Sechrest, & Reno, 1990). As noted by Onwuegbuzie and Daniel (2000):

Unfortunately, some researchers have found that the majority of published studies and dissertations are seriously flawed, containing analytical and interpretational errors (Daniel, 1998a; Hall, Ward, & Comer, 1988; Keselman et al., 1998; Onwuegbuzie, 1999a; Thompson, 1998a; Vockell & Asher, 1974; Ward, Hall, & Schramm, 1975; Witta & Daniel, 1998). Some of these flaws have arisen from graduate-level instruction in which research methodology and statistics are taught as a series of routine steps, rather than as a holistic, reflective, integrative process (Kerlinger, 1960; Newman & Benz, 1998); from graduate-level curricula that minimize students’ exposure to quantitative and qualitative content (Thompson, 1998a); from proliferations of various erroneous “mythologies” about the nature of research (Daniel, 1997; Kerlinger, 1960); from increasing numbers of research methodology instructors teaching out of their specialty areas; and from a failure, unwillingness, or even refusal to recognize that analytical and interpretational techniques that were popular in previous decades no longer reflect best practices and, moreover, may now be deemed inappropriate, invalid, or obsolete (Schmidt & Hunter, 1997). (pp. 3-4)

Examples of inappropriate statistical practices that are still being disseminated in many statistics courses include (a) not providing evidence that statistical assumptions were checked prior to conducting inferential analyses; (b) not discussing power/sample size considerations; (c) inappropriate treatment of multivariate data; (d) use of stepwise
procedures; (e) failure to report reliability indices for either previous or present samples; (f) no control for Type I error rate; and (g) failure to report effect sizes (Onwuegbuzie & Daniel, 2000, in press).

An important decision made by statistics teachers relates to the emphasis that should be placed on measurement. Aspects that fall under the auspices of measurement include classical test theory and item response theory. Clearly, the level of statistics course, number of statistics courses in the sequence, level of students, type of student body, and other attributes presented in Figure 1 largely determine the amount of exposure to measurement issues that students will receive. For example, in an introductory-level course, it is likely that time will prevail to teach only the basics of classical test theory, such as the concepts or reliability and validity.

Similarly, the role of qualitative research is another decision that a statistics instructor must make. For example, at some institutions such as Valdosta State University, doctoral students are required to complete a course in mixed methodologies, in which they are taught how to conduct research that utilizes both quantitative and qualitative data either in a parallel or sequential manner (Onwuegbuzie, 2000b). The likelihood of statistics instructors using such a pragmatist approach to teaching statistics is a function of their philosophical orientation (i.e., world view), as well as their experience in using mixed methodologies.

Other questions that statistics educators should address are (a) What is the role of program evaluation and research design in statistics classes? (b) Should statistics instructors concentrate more on theory or application? (c) What is the role of computers in statistics classes? (d) Should statistics instructors focus on computational formulae? What should the ratio be between descriptive and inferential statistics? (e) What should
be the balance between hypothesis testing and estimation? (f) Should Bayesian
statistics play a role in statistics classes? (g) What should be the balance between
graphical and analytical techniques? (h) What is the role of meta analyses? (i) How
much can we expect reasonably for students to learn within one statistics course and
within a series of statistics courses? and (j) What should be the role of action research?

The American Psychological Association (APA) Board of Scientific Affairs, who
convened a committee called the Task Force on Statistical Inference, provided
recommendations for the use of statistical methods (Wilkinson & the Task Force on
Statistical Inference, 1999). Useful recommendations were furnished by the Task Force in
the areas of design, population, sample, assignment, measurement, results, analysis, and
discussion. At the same time, the Committee on Professional Ethics of the American
Statistical Association (ASA) addressed the following eight general topic areas relating to
ethical guidelines for statistical practice: (a) professionalism; (b) responsibilities for funders,
clients, and employers; (c) responsibilities in publications and testimony; (d) responsiblities
to research subjects; (e) responsibilities to research team colleagues; (f) responsibilities to
other statisticians or statistical practitioners; (g) responsibilities regarding allegations of
misconduct; and (h) responsibilities of employers, including organizations, individuals,
attorneys, or other clients utilizing statistical practitioners (The American Statistical
Association, 1999, p. 4). Thus, statistics instructors should consider designing their courses
with respect to both the APA and ASA recommendations.

Finally, the debate concerning statistical significance testing (e.g., Daniel, 1998b,
Ernest, 1998; Nix & Barnette, 1998a, 1998b; Thompson, 1998b) makes it clear not only
that statistics is a controversial subject matter (Derry, Levin, & Schuuble, 1995), but also
that it represents an art rather than a science. As such, as recommended by Derry et al. (1995), statistics “should not be taught as a set of final-form, universally accepted concepts that can be handed down by authority and conveyed to students by teachers and textbooks” (p. 52). Moreover, statistics courses should serve to make students cognizant of the major statistical debates (Derry et al., 1995; Johnson, 2001). According to Derry et al. (1995), “students could gain even more by actually discovering and participating in such controversies” (p. 52). Chance (1997) echoes these sentiments, stating that students’ most meaningful learning gains arise from debating ideas form each other.

**Pedagogical Style in Statistics Courses**

In addition to considering the context and content of statistics courses, instructors must focus on the pedagogical issues. An important aspect of this is how to deliver statistical knowledge (i.e., the media). Central to the delivery system is the role of web-based instruction (i.e., synchronous and asynchronous) in statistics classes. Web-based instruction typically involves the use of computer software and hardware that do not represent the mainstream utilized by college instructors. As such, statistics instructors who use such delivery systems must have an interest in technology, a willingness to use technology, knowledge of various computer software, and a willingness to change. Additionally, the likelihood that this teaching mode will be employed will be greatly facilitated if the statistics teacher should have prior experience of using technology as a teaching aid, as well as confidence in using technology and relatively low levels of technology-related anxiety. Figure 4 presents all these variables in a conceptual model. Also presented in Figure 4 are time, resources, and administrator support. Obviously these three components must be in place for delivery systems such as web-
based teaching to prevail. The final ingredients in the conceptual model are instructors’
learning style and teaching style. For example, if an instructor’s learning style is suited
to a web-based learning modality, then he/she is more likely to teach using this style.
Interestingly, Onwuegbuzie and Daley (1997) found that students who are most similar in
learning style to their instructor with respect to persistence orientation, peer orientation,
auditory preference, and multiple perceptual preferences tend to obtain higher levels of
performance in educational research courses.

Within the statistics classroom, statistics teachers should experiment with
different organizational approaches. For example, cooperative learning techniques
(Collins & Onwuegbuzie, 2001; Onwuegbuzie, in press-b; Onwuegbuzie & DaRos,
2001; and advance organizers (Onwuegbuzie, 1999b) could be examined. With respect
to the former, Derry, Levin, Osana, Jones, and Peterson (2000) found that a course in
which most instruction was anchored to mentored, small-group collaborative activities
that simulated complex, real-life problem solving led to meaningful gains in students’
reasoning ability. With respect to the latter, Onwuegbuzie (1999b) found that students
enrolled in the advance organizer sections of a research methodology course obtained
higher levels (Cohen’s d effect size = .54) of overall achievement than did their
counterparts enrolled in sections in which advance organizers were not utilized by the
instructor.

As noted by Johnson (2001, p. 2), statistics teachers “must help students
become effective users of computer packages (e.g., SPSS, SAS, DataDesk)…[and]
because of the widespread availability of statistical software…[statistics instructors] can
now focus much of [their] effort on teaching students how to correctly use statistical
software and how to interpret output.” In order to meet the technological needs of
students, Valdosta State University required that every statistics and research methodology course take place in a computer laboratory, in which every student enrolled in these classes sits at a computer terminal. (The computer monitor is built into the workstation such that the student’s view of the instructor is not impeded.) In these laboratories, the instructor has a computer console positioned at the front of the class containing a computer linked to a projector. Students enrolled in statistics and research methodology courses are then able to learn how to use computer software (e.g., SPSS) in a hands-on, step-by-step manner. Instructors at Valdosta State University believe this to be an effective way of teaching students how to analyze real data.

A Typology of Assessments of Statistics Learning

Assessing statistics learning represents the final element in the quartet of considerations pertaining to statistics courses. As mentioned above, 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.

As noted by Garfield (1994), the fundamental decisions that the statistics teacher has to make in the area of assessment can be classified into 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,
Assessment As a Tool for Learning Statistics

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 learn how to evaluate their own knowledge, skills, and applications of their 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 include 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 (2000a), 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.

Assessments that have been used in statistics courses include quizzes, in-class
examinations, take-home examinations, term projects with peer reviews and oral
expectation, 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, 2000a).

In-class examinations, which are the most traditional in nature, 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 (2000a) 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.

A problem with most statistics courses 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). These educators define statistical authenticity 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 activity dimension pertains to “the extent to
Assessment As a Tool for Learning Statistics

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’
assessments are utilized, 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 (2000a, 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) 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,
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 are used (Wilson & Onwuegbuzie, 1999).

In undertaking the mini-dissertation, students are encouraged to utilize archival data because of the relatively short time required (i.e., one semester) to complete their study. Advantages of utilizing archival data sets are that (a) “data sets like these give students a taste of ‘big time’ statistics; (b) “many of the data sets have a complex structure and are rich with statistically interesting features”; and (c) “materials to support particular curricular uses can be prepared in advance” (Cobb, 1993, paragraph 33-35).

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 a journal-ready article 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.

Interestingly, Onwuegbuzie (2000a) 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 is perceived by many statistics instructors to be inadequate (Curtis & Harwell,
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 extensive information to statistics instructors and others about assessments in statistics courses. In particular, a typology was provided of different effective ways of assessing performance in statistics classes for the various contexts (e.g., undergraduate vs. master’s vs. doctoral), content (e.g., measurement vs. evaluation vs. research design), and pedagogical styles (e.g., web-based vs. traditional; theory vs. concept vs. application).

Students prepare for examinations in ways that reflect how they believe they will be tested (Crooks, 1998; 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 enter 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, 2000a). Finally, 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, symposiums, 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.
Assessment As a Tool for Learning Statistics

References


treatment interactions and Matthew effects. Paper presented at the annual conference of the American Educational Research Association (AERA), Seattle, WA.


Levin, J.R. (1998). What if there were no more bickering about statistical significance tests? *Research in the Schools, 5*, 43-54.


Assessment As a Tool for Learning Statistics

Educational Research Association, San Diego, CA.


Onwuegbuzie, A.J., & Daniel, L.G. (In press). *Uses and misuses of the*
correlation coefficient. Paper presented at the annual meeting of the Mid-South Educational Research Association, Point Clear, AL.


Assessment As a Tool for Learning Statistics


Figure Caption

Figure 1. Model of teaching and learning statistics.
Assessment As a Tool for Learning Statistics

Context of Teaching Statistics

Content of Statistics Classes

Assessing Statistics Learning

Pedagogical Style
Figure Caption

*Figure 2.* Considerations regarding the context of teaching statistics.
Figure Caption

*Figure 3.* Factors mediating content of statistics courses.
Figure Caption

*Figure 4.* Factors mediating pedagogical style in statistics courses.