

Rethinking Assessment of Student Learning in Statistics Courses

Joan GARFIELD, Andrew ZIEFFLER, Daniel KAPLAN, George W. COBB,
Beth L. CHANCE, and John P. HOLCOMB

Although much attention has been paid to issues around student assessment, for most introductory statistics courses few changes have taken place in the ways students are assessed. The assessment literature describes three foundational elements—cognition, observation, and interpretation—that comprise an “assessment triangle” underlying all assessments. However, most instructors focus primarily on the second component: tasks that are used to produce grades. This article focuses on three sections written by leading statistics educators who describe some innovative and even provocative approaches to rethinking student assessment in statistics classes.

KEY WORDS: Fairness; Feedback; Grading; Statistics education; Testing.

1. INTRODUCTION (JOAN GARFIELD)

In statistics education, there have been conferences, workshops, and publications on various aspects of the assessment challenge (see Garfield 1994; Gal and Garfield 1997; Chance 2000, 2004; Weldon and Phillips 2007). The ASA-endorsed Guidelines for Assessment and Instruction in Statistics Education (GAISE; American Statistical Association 2005) include the following recommendation: Use assessments to improve and evaluate student learning. Many of these sources emphasize that students will value (and focus their learning on) what their instructors assess.

Despite the prominence of assessment in the GAISE recommendations and in the statistics education literature, few statistics faculty have taken time to think about the bigger issues involving student assessment or have created new approaches to this endeavor. While many instructors think carefully about

how to write a good quiz or exam or how to score a student project, there are bigger issues that need to be considered as well.

1.1 Overview

An invited session offered by the ASA Section on Statistics Education at the 2008 Joint Statistical Meetings presented three creative, and arguably controversial, approaches to rethinking the challenges of student assessment. These presentations, now sections in this article, offer a refreshing alternative to traditional ways of thinking about assessment only in terms of testing and grading. The goal of this article is to spur further discussion and debate.

While George Cobb (Section 4) proposes a system that goes against traditional measurement principles of “fairness” in order to promote student effort and to individually maximize each student’s successful learning, Daniel Kaplan (Section 3) applies a statistical approach to modeling student learning in a class and challenges us to apply our statistical methods to the analysis of student data in our courses. Andrew Zieffler (Section 2) provides a model of assessment FOR learning and assessment AS learning, in addition to the more traditional view of assessment OF learning. His discussion, based on research in the growing field of student assessment, as well as Kaplan’s, challenge instructors to examine each individual item in their assessment plan as part of a larger whole to gather greater information and to ensure the alignment of the overall assessment program with course goals.

Together, these three authors challenge us to rethink many of our assumptions and practices regarding student assessment in statistics courses. The final section, written by John Holcomb and Beth Chance, responds to these three authors and provides both a critique and consideration of practical issues for implementation.

1.2 Questions Every Statistics Instructor Should Consider

Although the ideas presented in this article may seem extreme or controversial, it is worth considering the issues raised and then rethinking one’s own assessment practices in light of the following questions:

1. How well do my assessments motivate students to do their best and reveal what they know?
2. How fair are my assessment practices to different types of students?
3. How well do my assessments capture and foster the most important learning outcomes that I value?

Joan Garfield is Professor, Educational Psychology, University of Minnesota, Minneapolis, MN 55455 (E-mail: jbg@umn.edu). Andrew Zieffler is Lecturer, Educational Psychology, University of Minnesota, Minneapolis, MN 55455. Daniel Kaplan is DeWitt Wallace Professor, Department of Mathematics, Statistics, and Computer Science, Macalester College, Saint Paul, MN 55105. George W. Cobb is Robert L. Rooke Professor of Statistics, Emeritus, Mount Holyoke College, South Hadley, MA 01075 and Vernon Wilson Endowed Visiting Professor, Department of Mathematics and Statistics, Eastern Kentucky University, Richmond, KY 40475. Beth L. Chance is Professor, Department of Statistics, California Polytechnic State University, San Luis Obispo, CA 93407-0405. John P. Holcomb, Jr. is Ph.D., Professor and Chair, Department of Mathematics, Cleveland State University, Cleveland, OH 44115-2214.

4. How well do my assessments inform students of their learning, and provide useful feedback to them to improve their learning?
5. Am I relying on methods that are easy to use (e.g., test banks from publishers) or am I really using the assessments that do the job best? (And what do I mean by “best”?)
6. How satisfied am I with my method of assigning grades?
7. What needs to change in my assessment practice?
8. How well do my assessments promote learning?

It is time to assess our own assessment practices and instruments and to critically evaluate their relationship to the important, desired student outcomes. I hope the discussions in this article succeed in stimulating that endeavor.

2. MAKING STUDENT LEARNING THE PRIORITY OF ASSESSMENT (ANDREW ZIEFFLER)

To the many teachers of statistics who have managed to spend a career in innocence of formal research in education, the word “assessment” is sometimes misunderstood as no more than a fancy word for grading. Indeed, the next two sections of the article by Kaplan and Cobb, teachers of statistics without formal training in education research (despite their regard for that science), tend to emphasize the facet of assessment that deals with assigning grades. Nevertheless, within the education community, “assessment” refers to a much broader enterprise, one with multiple purposes. The present section offers statisticians an overview of those multiple purposes, with emphasis on learning, rather than on evaluating what has been learned. In addition, the section provides both a brief introduction to assessment research and practical examples.

2.1 Assessment *For* Learning

Frequently, assessments used by classroom teachers (e.g., homework, quizzes, exams, etc.) are used to assign grades to students—assessment *of* learning. These grades provide a concise snapshot of their performance or learning at a given point in time. The educational measurement literature has provided much guidance for the creation and use of this type of assessment (e.g., Arter and McTighe 2001; Haladyna, Downing, and Rodriguez 2002; Rodriguez 2005; Downing 2006; Schmeiser and Welch 2006).

An important potential use of assessment is to explicitly include it as a part of the learning process—assessment *for* learning. Effective assessment *for* learning can better inform teachers about the nature of student learning through the transparency in the learning cycle (Instructional goals–Assessment–Feedback). Inclusion of such items allows students to engage in learning via the assessment process by allowing students to communicate their thinking about the content and process in relation to what is expected (Atkin, Black, and Coffey 2001; Harlen 2006). This in turn allows instructors to provide clear and instructionally meaningful feedback to students, which promotes better learning (e.g., National Research Council 2001; Stiggins 2002; Bass and Glaser 2004; Gardner 2006). This all requires a coherent and explicit relationship between the classroom assessments, curriculum, and instruction.

2.2 Alignment

Alignment refers to a match between the assessments being used and the curriculum being taught. Alignment involves not just the content of a course, but also the level of performance, emphasis and depth of material, and consistency with performance standards. Alignment is an important goal. Research has suggested that an aligned system of classroom instruction and assessment leads to better student performance (e.g., Gamoran et al. 1997; Rowan 1998). The synchrony between assessment and instruction also reinforces expectations for teaching and learning.

Alignment is best achieved through a deliberate process. One of the simplest methods is to create a two-way table (i.e., instructional blueprint or content matrix) where one dimension represents the content domain that is covered in the course, and the other dimension represents the desired cognitive demand or expectation (e.g., Porter et al. 2007). After the instructional blueprint has been created, each individual assessment item can be placed in an appropriate cell. The blueprint can be analyzed to determine what is being assessed and at which level (e.g., Bloom’s Taxonomy; see Bloom 1956). The instructor can then develop assessment items to fill in underpopulated cells.

2.3 Classroom Formative Assessment

An underutilized type of assessment in many statistics courses is *classroom formative assessment*. Classroom formative assessments (e.g., having students write a “minute paper” in which they respond to a question that elicits what they comprehend or do not comprehend prior to leaving class) are designed so that instructors can provide specific information to students about their understandings and misunderstandings *during* the learning process. Because they are designed with this intent, classroom formative assessment supports, enhances, and improves learning by often informing the student through written feedback rather than a score or grade.

In a review of research from many different countries (over 250 articles), Black and Wiliam (1998a) found that an intentional focus on formative assessment led to unprecedented positive effects on student achievement. They also found that the effects were even more substantial for low performing students’ writing: “improved formative assessment helps low achievers more than other students and so reduces the range of achievement while raising achievement overall” (Black and Wiliam 1998b, p. 141).

Formative assessment facilitates the learning cycle (Instructional goals–Assessment–Feedback) by allowing students to communicate their thinking about the content and process in relation to what is expected (Atkin, Black, and Coffey 2001; Harlen 2006). This conveys their understandings and misunderstandings to the teacher and allows for modification of instruction if necessary, which can immediately improve the learning process. Examples of formative assessment in the statistics classroom can be found in the works of Jordan (2004) and Mosteller (1989). More general discussion can be found in the book by Angelo and Cross (1993).

2.4 The Content of Feedback

Students can learn from the feedback that teachers give on assessments: formative feedback. The literature on feedback indicates that the nature and quality of the teacher's comments are some of the most important factors in the effectiveness of the feedback (e.g., McCloskey and Leary 1985; Butler 1987). While there are many recommendations based on the literature for delivering feedback (see Shute 2008 for a more extensive list), there are a few that seem particularly salient: focusing on specific features of the student's work in relation to the instructional goals, giving clear suggestions for how to improve in progressing toward those goals (e.g., Corbett and Anderson 2001; Narciss and Huth 2004), helping students identify their strengths and weaknesses without an overall grade attached (see Wiliam 2007), and using praise sparingly, if at all. Research has suggested that the use of "praise" distracts students from elements of the feedback that are important for improving their work and thus inhibits overall learning (e.g., Kluger and DeNisi 1996).

Sadler (1989) has suggested that it is insufficient for teachers to only provide feedback about whether answers are right or wrong if the feedback is to facilitate learning; feedback should be linked explicitly to the learning goals and should suggest strategies for improvement. Providing specific details in the feedback to students about how to improve their answers can alleviate student frustration and uncertainty about how to respond to the feedback given (Fedor 1991; Williams 1997). Shute (2008) added that effective and useful feedback should be timely, providing "real-time" information as the learning is occurring, so that the student can respond productively to it rather than waiting until after an exam when many students view it as too late.

2.5 Collaborative Assessment

Collaboration among students can enhance learning. (For a comprehensive summary of this literature, see Roseth, Johnson, and Johnson 2008.) When assessment is aligned with classroom experiences, assessment itself can be an opportunity for student collaboration. It is important, however, that all group members understand the concepts and topics addressed in the assessment. The process of collaborative assessment in statistics education is well described by Roseth, Garfield, and Ben-Zvi (2008), who also provided many practical examples of how collaboration can be used in a statistics classroom, for example cooperative quizzes and group projects.

2.6 A Classroom Assessment System

Classroom assessment functions within the larger integrated structure of curriculum and instruction. For optimal student learning, each component should work in synchrony to enable students to meet the learning goals. Alignment of assessments, curriculum, and instruction in the statistics classroom helps set the stage for meaningful and effective student learning to take place. Furthermore, the use of classroom formative assessment, collaborative assessment, and the provision of appropriate feedback to students can accommodate and supplement the learning process.

Assessment *for* learning provides benefits to both students and instructors alike. Students benefit from the formative feedback and communication that allow them to monitor their progress and focus their efforts on the weaknesses and misunderstandings identified by assessment. By directing their own learning based on the feedback from assessment, they can gain the tools needed to set the foundation for a lifetime of learning (Stiggins 2002). Teachers benefit from a good assessment system by enabling themselves to make better instructional and grading decisions, informed by assessment evidence that accurately reflects students' progress toward desired learning goals.

3. USING STATISTICAL TOOLS AND TECHNOLOGY TO GRADE STUDENTS (DANIEL KAPLAN)

To many faculty (if not all), grading is to teaching as carbon emission is to manufacturing—an unwanted byproduct, best minimized. This section first summarizes some of the standard compromises teachers make when grading, then shows how technology offers new options for assessment and argues that, as statisticians, we should apply the methods of our profession to make assessment simultaneously more motivating for learning and more accurate for grading.

3.1 Assessment Goals and Challenges

The goals of grading have been cleverly summarized as assessment *of* learning, assessment *for* learning, and assessment *as* learning (Western and Northern Canadian Protocol for Collaboration in Education 2006). To be more explicit, consider these statements of the three "of/for/as" goals:

- *Evaluating the individual student.* This is "summative assessment," giving a grade.
- *Assessing the teaching.* How effectively are the content, concepts, and methods being taught? Are the students learning? Is the class time being used effectively? Are the topics too advanced or too elementary for these students? Sometimes this is called "formative assessment" and relates to adopting the teaching to meet the students' needs.
- *Motivating and guiding the student.* Grading students (and providing direct feedback) helps motivate students to keep up with the class and do the work needed to learn.

To be honest, we ought to acknowledge some other goals that also shape how we grade.

- *Making the best use of our time.* As the saying goes, "I teach for free; they pay me to grade." Few instructors enjoy grading and most have better things to do with their time. As such, instructors reasonably try to be efficient when grading.
- *Organizing our classroom time.* Instructors often spend class time reviewing homework exercises. Many instructors use in-class activities, often in the form of group work, which are also a form of assessment and feedback.
- *Presenting a professional, dignified, and rich view of statistics.* This can mean, to some people, posing challenging problems. It can also refer to the use of genuine data in realistic settings and problems analogous to work encountered in a professional setting, so called "authentic" assessment.

In some ways, these various goals can support and reinforce one another. The prospect of the end-of-term exam is a powerful motivator to many students to get their work done. Use of “authentic” problems can pique student interest and, by showing the practical importance of statistical techniques and concepts, help to inspire students to study. Information from the formative and summative evaluation of students can provide evidence on the effectiveness or ineffectiveness of the teaching.

The ways in which the goals conflict are perhaps less obvious, in part because we adopt grading practices that avoid areas of conflict. Consider, for example, group work. It is believed that working and studying in groups can help students in a variety of ways. Yet group work can be problematic when it comes to using the information for evaluation of individual students. On some assignments, many instructors regard collaborative work as a form of cheating and simply forbid it.

It is extremely tedious to score activities that are intended to drill students in specific tasks. But drill is an accepted technique in many areas. When we avoid drill, is it because we think it is unproductive, or because it conflicts with our goals of making the best use of our time or of presenting a rich view of statistics?

For the purposes of assessment, it can be helpful to ask students questions that the instructor assumes are too easy or are too hard. Such questions give an idea about the range of student knowledge and ability. Asking questions about material that has not yet been covered in class can reveal areas where the students are capable of learning the material on their own and can focus students on the topics that will be treated in class so that they come prepared with questions. Yet this conflicts with traditional views of evaluation since students are put off by such questions, particularly ones that are too hard. They regard it as unfair to be evaluated on material that has not yet been covered. Students are also disturbed when they receive low scores, as they will on questions that are too difficult.

At the same time, instructors refrain from asking questions that are too easy, even though student answers to such questions can be important for assessment, pointing as they may to the need for remedial work or to allowing instructors in the future to skip over material that has already been mastered. Questions that are too easy conflict with the goals of providing a rich view of statistics and of evaluation and motivation since students can get high grades by answering easy questions with little work.

3.2 Traditional Assessment Practices

Instructors often adopt grading practices that avoid conflicts between the various goals, but these same practices may also limit their ability to achieve those goals. Some common practices of teachers include the following:

- *Score rich, “authentic” questions.* Avoid drill. It can be tedious to score drill questions, unless the process is automated. Drill is also seen as too easy to provide meaningful evaluation.
- *Collect assignments in class, then review them.* Do not allow students to revise their work after it is handed in. Why? It takes a long time to collate the assignments that students hand in. If students can revise their work after the instructor has gone over a problem in class, the assignment does not work very well for evaluation.

- *Record a single score for each assignment, summarizing many individual items.* Sometimes, instructors have just a few graded assignments during the term. This simplifies book-keeping and reduces the amount of time spent shuffling papers. Unfortunately, it also means there are no data about student performance on individual questions and little opportunity for the instructor to respond to problems during the term. Highly aggregated assignment scores do not support assessing teaching very well.
- *Record missing work as zero.* This motivates students to hand in the work, but means that students need to take the time to do work that may not be needed for them.

These conventional practices are rooted in the days when students wrote out their answers on paper and handed them in when they came to class, and when instructors kept a paper gradebook and computed final grades as simple sums. The advent of modern computing has led to a veneer of technology being laid over the traditional forms of grading. Students now use word processors to write their homework before printing it out. At some schools, the work might be handed in by email, uploaded into a course management system, or scanned in the department office. Instructors use Excel to keep their gradebooks, and use the sum or average macros or commands to perform the calculations that used to be done by hand.

3.3 Grading Statistics as Statisticians

Computer technology can enable us to avoid many of the trade-offs described above. In particular, computers can support scoring at a very fine level of detail, item by item. Computers also make it possible to provide very fast feedback so that students know when they are doing well and so that instructors can assess what concepts have been learned in a timely manner to guide their teaching and classroom activities.

However, it is not just computer technology that is needed. There is an intellectual technology required as well: the ability to organize data and to extract information from large amounts of data, the ability to handle missing data, the ability to distinguish between individual student performance and teaching effectiveness. As it happens, statisticians already possess this intellectual technology.

Imagine if educators had at hand a system for writing and sharing assessment items that supported students submitting answers online, ranging from simple forced-choice questions to complete, free-text essays. Those items that can be graded automatically will be, with the instructor or graders scoring free-text or other complex answer forms. The system would remove the overhead of collecting and returning papers and entering scores.

With such a system in place, instructors could adopt new practices to reduce the conflicts among the various goals of grading.

- *Grade at a very fine level of detail, item by item.* Let computers do the work of organizing the data; instructors should be exercising judgment, not doing bookkeeping. By grading submissions at an item-by-item level, instructors are able to probe student performance in a detailed way.

- *Do not avoid drill.* Give students credit for doing it, but do not require it. Making it easy for students to submit answers on-line allows time to be spent on drill, rather than handling paper.
- *Ask many questions that are too easy or too hard,* to help figure out what is the right level. Since items are graded at a fine level of detail, it is easy to identify which ones are too easy and which too hard.
- *Use modeling techniques* to cope with missing data and to adjust for the difficulty of the questions as revealed by the student answers. A simple average of each individual student's scores cannot do this; it will be distorted if questions are too easy or too hard or if there are missing data. But scores need not be simple averages.
- *Make assignments that are submitted and scored before class* so that the instructor can use the information in planning the class.

The electronic communications and database infrastructure needed is largely in place to implement the sorts of systems that would support such practices. We appear to be on the cusp of a major change from publishing on paper to distributing in electronic form, as evidenced by the success of on-line readers such as the Kindle and iPad.

A major obstacle, however, is the lack of an easily used standard for writing and sharing materials that support on-line work. Many publishers use proprietary systems that do not interoperate well, making it difficult to adopt new materials from different authors or to integrate materials from different sources. To make the point by way of a contrast, consider how instructors and authors routinely share content: document files, images, data, and computer software. But what are the standard formats for sharing materials that fit into an electronic scoring system?

It is unclear how to bring about the development of standard and easy-to-use formats for on-line grading. Any system requires both authoring tools to prepare the materials and servers for collecting student answers and providing ready access to them. The system has to be simple enough that instructors can make use of them without an undue amount of training. To demonstrate how a system might work, we have developed a prototype on-line system, AcroScore, that makes use of simple LaTeX markup and produces documents that connect with a small server that can be installed locally. The system uses standard URL coding and therefore can be adapted to just about any modern word processing system.

Once a suitable data collection system is in place, there is still a need for appropriate data analysis. Statisticians are uniquely positioned to develop ways of combining item-level scores into an overall end-of-term grade and that support the various goals of grading. One effective approach is to fit models to the item-by-item score data: score modeled by student and item. The coefficients on the different levels of student give information about the relative rank of each student within the class. The coefficients on the items tell how difficult or easy that item was based on the student responses. In deciding on final grades for each student, I use several related models:

- An “achievement model” where missing scores are not considered.
- A “work product model” where missing scores are treated as zeros. This rewards students who complete large numbers of items.
- An “outlier-free model” which is the achievement model with low outlying items discarded for each student. The items to be discarded are those where the residual is negative and large in magnitude.

There are several advantages to such an approach. It is easy to deal with those questions that are valuable for assessment but are too easy or too hard. Instructors get information about how easy or hard individual items were. Missing data can be handled without being inappropriately punitive. It also illustrates for students a direct, meaningful application of the statistical content they are learning to their own lives.

As the ongoing revolution in publishing unfolds, and as we deploy systems that exploit the potential of web communications, we should be mindful that the grading techniques that are so widely used evolved in a very different technological environment and are not optimally suited to achieve our goals. Statisticians ought to be in the vanguard of designing the new systems to take advantage of our abilities to collect, analyze, and interpret complex data. Statisticians long ago moved beyond simple averages in the analysis of data. It is time to bring those skills and techniques to bear on how we evaluate our students and assess our teaching to guide us in improving both.

4. AGAINST FAIRNESS (GEORGE COBB)

Zieffler (Section 2) has focused on using assessment to foster learning—learning by students about course content and learning by teachers about teaching effectiveness. Kaplan (Section 3) has focused on ways to combine technological advances with statistical thinking to make grading less intrusive, more motivating, and more accurate. This section notes how a broadening recognition that data matters has been making classes increasingly heterogeneous, and argues that such variability in student backgrounds, interests, and talents behooves us to reexamine conventional notions of fairness in grading.

4.1 Context

Undergraduate statistics education is changing rapidly. Rapid enrollment growth in advanced placement courses in high school and in introductory courses in colleges is creating demand for new courses beyond the level of the first course. Often these courses must serve a variety of interests and backgrounds. For example, some students may have comparatively strong backgrounds in mathematics and an interest in underlying theory along with the applications; other students may have an interest mainly in statistics as a tool for work in sociology or psychology. It is my thesis that such a broad spectrum of backgrounds and interests poses an assessment challenge for those of us who teach.

My goal in this section is to raise three concerns about some traditional thinking about assigning grades. (The phrase “assigning grades” instead of assessment is meant to acknowledge that the criticisms I offer are directed against a narrow notion of assessment that many would now consider old fashioned.)

The approach that I propose as an alternative has obvious limitations. In particular, I cannot imagine how to make such an approach practical for a class of more than about twenty students. Nevertheless, I hope that many of the issues I raise will invite teachers to think in useful ways about their own approaches to assessment and grading, especially in statistics courses for students with heterogeneous backgrounds and interests.

4.2 Fairness

My first and main concern is fairness. . . I am against it. I am going to argue that a commitment to a misdirected notion of fairness is responsible for a variety of unintended evils in teaching and learning, and constitutes one of the main obstacles to authentic assessment. The bare bones argument is a simple one: Fairness involves comparing students with each other. Accepting fairness as a goal pushes us toward quantification, toward uniformity, and toward a view of learning as a zero-sum game.

If we quantify, it becomes easy, almost inevitable, to hold everyone to the same standard: Your grade is based on your score. This nominally “fair” approach is easy to defend, in the sense that you can explain why one student’s grade is higher or lower than another’s, but it is not easy to defend in terms of its consequences for learning.

Think about two sets of consequences, for students, and for curriculum.

Students. First, the consequences for students. Students vary; their backgrounds vary; their interests vary; their learning goals vary; their learning styles vary.

Focus for a moment on the so-called “stronger” student. If everyone is graded on the same scale, teachers tend to feel less responsibility to challenge the better-prepared students. Students who start with more background or who pick things up quickly can expect to do well even if they work a lot less than students who start with less background or who learn more slowly. This is just how it is most of the time. “Stronger” students are misled into learning less than they might, and learning suffers. Our system penalizes talented students who should be learning to stretch themselves, and encourages a tendency to aim low. For these students, fairness encourages laziness and low standards for learning goals.

Now consider the so-called “weaker” students. If the assessment system is fair in the nominal sense, these students start with the deck already stacked against them. Quite likely they think from the beginning that it is impossible for them to do well, and they are resigned to a lower grade no matter what they do. They also may feel pressure to try to learn things they may not be ready for yet, or to settle for a superficial kind of learning, in order to keep from falling farther behind. Here, fairness encourages low self-esteem, and low standards for what it means to understand.

For all students, both the more prepared and the less prepared, both the quicker learners and the slower learners, misdirected notions of fairness encourage a sense of competition, discourage helping others, and encourage students to judge themselves and their accomplishments by comparing themselves with others, rather than judging themselves by what works best for them as individuals.

Curriculum. Let us turn now from students to curriculum. Because our concern with fairness encourages uniformity of expected outcomes, our system tends to foster an artificially compartmentalized curriculum. I find it useful to think about learning simultaneously in terms of the process and its content, so, as in quantum theory, learning has a kind of wave/particle duality. Fairness pushes things too far in the direction of the particles.

Consider the matter of prerequisites. Misdirected attention to equity says that we should measure all students using the same yardstick at the end of the semester. If everyone is to be graded on the same scale at the end, the system is fair only if we assume that everyone starts out at roughly the same place, which is what our grading system encourages us to expect of students. This means we are supposed to have a fairly strict prerequisite structure. Everybody has to learn *abc* before the teacher is allowed to talk about *def*. You are supposed to take *this* course first, *that* course second.

Uniformity encourages us to deliver properly compartmentalized and sequenced quanta of knowledge; courses become standardized bags of curricular McNuggets; our curriculum becomes a succession of intellectual Happy Meals.

An alternative wave theory sees learning as a process. The best learning has a narrative flow, with a beginning, development, resolution, and integration. The learning occurs as our students weave the threads of new ideas into their existing understanding. Within reasonable limits, it should not matter if students do not all start from the same place. Two different students taking the same course will inevitably get different things from it. We should embrace that inevitable difference, and try to see that each student gets as much as possible from our course, regardless of starting place.

Alternative. Instead of distorting assessment for the sake of fairness and uniformity, we can look to an alternative model in the form of the one-room school, with students in several different grades learning different things in parallel. Generations of teachers have made that model work. Why can we not think of each of our own classes as a more uniform version of the one-room school, with several levels together in a single room? The parallel is far from exact, of course, but as a metaphor for an alternative to what our system would have most college teachers do now, it can be useful. To make it work, we have to offer more options, for example, *more* homework, and more *kinds* of homework, than we expect any one student to do. Some will spend more time on drill-and-practice; others will be able to skip most of that, in order to spend more time on open-ended investigations and challenge problems. We will need to encourage students to make their own choices from the work we offer, and we will need to reassure them that no one is expected to do everything, and that we expect *them* to decide for themselves which things to skip and which things to do.

In short, then, we should *abandon fairness*. Instead, we should *assess students as individuals*, in the context of their individual backgrounds, interests, goals, and learning styles.

4.3 Grade Inflation

The second of my three concerns is grade inflation. . . I am in favor of it. I am going to argue that if we take authentic as-

assessment seriously, then grade inflation is not just a likely consequence; rather, it is more: if our assessment is truly effective, then grade inflation should be inevitable.

Learning, done right, is exciting, and gives students a sense of accomplishment. Our main goal as teachers is to help ensure this experience for our students, and a good approach to assessment should serve this goal.

What gets in the way? For many students, the obstacles are things like feeling bored, feeling competitive, feeling unfairly judged, feeling not up to the task, feeling the kind of time pressure that keeps you constantly in catch-up mode.

If we assess well, we will help our students to see clearly what their accomplishments are, and help them to enjoy the satisfaction of a clear-eyed recognition of their own progress, regardless of where they start from. Good assessment will help us to help them direct their efforts as effectively as possible, and they will accomplish more, for the same effort. Their success will motivate them to work harder, and they will learn more as a result. If they are learning more, their grades ought to go up.

I used to think that I should judge outcomes, not effort. Students were wrong to feel that hard work alone was enough to entitle them to an A. If their final result was not up to the standards of the best in the class, their effort must not have been worthy enough.

I now give a lot more A's than I used to. I tell my students that if they and I can do a good job of directing their efforts, then how much they learn will depend mainly on how hard they work, and that everyone should be able to earn an A by putting in enough effort. In a typical class, although I try to be careful that students do not work so hard that they cheat their other courses of time, on evaluations almost all students describe their efforts in my course as either strenuous or fairly serious, with most saying strenuous, and the vast majority say that their enthusiasm for the subject has gone up. Students work harder, and enjoy it more, than they did back when I gave fewer A's.

So I urge you: *embrace grade inflation*. We need not guarantee every student an A, and I do not think we should do that, but we should grade in a way that allows every student to regard an A as truly within reach, and we should assess in a way that allows both us and our students to feel that the A's we give have been truly earned, because our assessment has made both the learning and the connection to effort easy to see.

4.4 Roger

My third concern is the one I have called Roger, based on my favorite cartoon. It is by Gary Larson, and the caption is "Roger screws up." It is my favorite because it captures the essence of my inner life. Roger, the guy who screws up, is the percussionist in a symphony orchestra. In the cartoon, he stands poised behind three rows of string players, waiting to sound the triumphal crash of cymbals that is supposed to provide the climax that the whole orchestra is building to. Acutely conscious of the weight of responsibility he bears, Roger has already raised his hands in readiness, and the balloon over his head shows him chanting to himself, "I won't screw up. . . I won't screw up. . . I won't screw up." His facial expression is one of extreme concentration, and you can just imagine him counting the measures until his big

moment, when, with all eyes upon him, he will bring his hands violently together. Although Roger does not know it yet, he is destined to blow it. A look at his hands reveals the disaster to come: Roger's left hand clasps a cymbal, but poor Roger's right hand is empty. There will be no triumphal crash. Despite his concentrated determination, Roger will screw up.

Authenticity. What does Roger have to do with assessment? I am going to answer my question with another question: "Which is better, to be sincere, or to be authentic?" Casually, we might be inclined to think of "sincere" and "authentic" as roughly synonymous, but a look at their origins suggests otherwise. I am going to argue that, in the original senses of these words, Roger is so preoccupied with being sincere that he has sacrificed his authenticity, and that, more generally, a misdirected sense of fairness in assessment encourages our students to make the same sort of unfortunate sacrifice.

At one time in the history of our language, to be sincere meant to be free from blemish, literally, "without stain." ("I won't screw up. . . I won't screw up.") Sincerity, in this older meaning, is fundamentally a conservative position, based on a minimax strategy: avoid blunders. Roger is so determined not to make a wrong noise at the wrong time that he ends up making no noise at all.

In his effort to be sincere, Roger has lost his authenticity. To be authentic is to be one's own agent, the author of one's deeds.

The tension between sincerity and authenticity parallels the tension between false fairness and effective assessment. Preoccupation with fairness leads to concern with grades, and ultimately, grades tend to be based mainly on "What did you get wrong?" Fair grading systems reward sincerity: being without stain. Authentic assessment is based on "What have you done well?" Authentic assessment should foster our students' sense of agency.

Hermit crabs. Intellectually, we—teachers and students both—are all a bit like hermit crabs. In order to grow beyond our current capacities, we have to abandon the protective shell of attitudes, skills, and old habits of thought that we have become comfortable with, but that now confine us. Abandoning that shell leaves us temporarily vulnerable—in particular, vulnerable to the embarrassment of screwing up—as we develop a larger way of thinking. If we want to encourage our students to grow, we have to make it safe for them to abandon their shells, safe for them to screw up. Our false notion of fairness works against creating the needed sense of safety. We substitute sincerity—being without stain—for authenticity—being one's own agent.

Consequences. What does this mean in practice? Two things, I think. First, it means that our assessment should acknowledge learning as a process that involves feelings and people. One thing I do in this connection is ask students to write weekly journal entries in which they talk about their learning experience—their goals, satisfactions, frustrations, accomplishments, and so on.

Second, it means that our assessment should emphasize the positive. We should be sure to point out what a student has done well. Emphasizing the positive does not mean being a Pollyanna, however. It does mean that, rather than making an issue of everything a student has done wrong, we should for the

most part focus attention on the few things we consider most important for the student to revise, and then make sure to provide an opportunity for revision.

I also avoid putting grades on written work. My premise is that if I can help a student direct her effort effectively, then how much she learns will depend on how much effort she puts in. In my comments, I try mainly to help her choose appropriate tasks and goals, I suggest revisions and extensions to the work she turns in, and I monitor effort, but in most cases I find that I do not have to comment on effort unless I think a student is working too hard.

4.5 Consequences

To summarize, then, here is a short version, in three parts, of what I think I have learned about assessment, the three things that I regard as most important.

- *First, abandon fairness.* Try to assess a student's work in the context of that individual's own background, interests, goals, and learning styles.
- *Second, embrace grade inflation.* Emphasize effort and accomplishments. If good assessment leads to better learning, students deserve higher grades. Low grades are often mainly a symptom of the shortcomings of our system.
- *Third, surprise Roger.* Make it safe to screw up. Emphasize things done well and things that a student can fix, and provide the opportunity for revision. Value authenticity over sincerity.

Assessment, I suggest, is ultimately just a systematic way of paying attention to our students. This is not at all to say that if we just pay attention, assessment will take care of itself. The *systematic* part is essential, and that is where the science and craft come in. That is where we need research and experts. But systematic does *not* mean and *should* not mean fair. Fair, no; systematic, yes. Sincere, no; authentic, yes. In short, authentic assessment should assess authenticity.

5. DISCUSSION (JOHN HOLCOMB AND BETH CHANCE)

In this discussion, we evaluate the approaches of Zieffler, Kaplan, and Cobb by adapting a framework developed under the auspices of the American Association for Higher Education (AAHE) Assessment Forum (Astin et al. 1992) aimed at defining good assessment practices. We outline how these principles relate directly to the ideas expressed by Zieffler, Kaplan, and Cobb, and we outline challenges for adopting the ideas they suggest. Lastly, we make suggestions for overcoming these challenges.

5.1 Principles of Good Assessment Practice

There is a large body of literature of assessment theory and practice, applicable across all disciplines. We find the eight principles of good practice for assessing student learning published by the AAHE (Astin et al. 1992) to be a particularly useful summary. These principles can be applied at the course level as well as to a broader program level of assessment. We use

the following modification of the principles as a framework for commenting on the approaches advocated by Zieffler, Kaplan, and Cobb.

1. Assess what you value.
2. Make the goals explicit.
3. Recognize that learning is multidimensional.
4. Pay attention to *process* as well as outcome.
5. Give timely feedback with indicators for change as well as reinforcement.
6. Develop buy-in from students and faculty and maintain feasibility of the assessment.
7. Allow for revision.
8. Make assessment authentic.

Zieffler's use of blueprints, formative assessment, and collaborative assessment model exemplary practice of assessment. The instructional and assessment blueprints provide excellent examples of making the process of "assessing what we value" explicit and achievable, especially helping the instructor identify the content and cognitive loads being demanded of students. Zieffler's uses of feedback through minute papers and in-class writing activities are effective ways to provide prompt, formative feedback in a supportive environment while also conveying to students that their input and understanding are important to the instructor.

Two main points of Kaplan's article involve the use of AcroScore for obtaining automatically scored responses from students regarding statistical questions that vary in level of difficulty, and then utilizing a modeling approach for calculating an overall grade based on scores from all individual items from materials throughout the term. Clearly, such a system of automatic scoring can be very beneficial in shortening the feedback loop, and tracking responses to individual questions will help instructors obtain detailed information about students' strengths and weaknesses as well as detailed evidence of how students come to learn statistical ideas or concepts. In addition, AcroScore provides quick analysis of weak points of collective student understanding that allow an instructor to change the content of the next class to address these needs ("Just in Time Assessment"). The AcroScore system also has tremendous potential for facilitating revision and sharing of items across instructors.

The major themes of Cobb's article involve individualized assessment based on progress that rewards effort and accomplishment. The author admits his approach is controversial and not implementable in all settings. Evaluating his approach in light of our guidelines shows that he assesses what he values—in particular, individual students' intellectual progress. In light of this individualized assessment, the instructor is able to capture the multidimensional nature of student understanding, while also enhancing students' ability for self-assessment. Cobb's use of journaling by students also allows him to assess the student's progress toward understanding as well as other components such as level of effort and thoughtful reflection. An individualized pace allows for multiple instances of feedback for students and revision by students. Cobb enhances the feedback mechanisms by making students feel "safe" to make errors and provides specific mechanisms for them to revise their understanding. Cobb also develops a classroom environment that allows

students to be authentic in the sense of being “one’s own agent, the author of one’s deeds.” This authenticity allows an instructor to reward achievement in regard to what a student has done well.

5.2 Challenges to Implementing Discussed Assessment Strategies

We note with some concern that Zieffler advocates a great many activities that take a great deal of time to create and administer. Anecdotal evidence suggests that students often forget many concepts and ideas six months after a class ends. A natural question arises, “Is such effort worth the pay-off for instructors?” When asked this very question, Zieffler responded that our ultimate goal as teachers is student understanding that lasts well beyond the final exam. He feels that the effort he has taken in his assessments does contribute to deeper understanding and better retention.

Kaplan’s use of the statistical package R, Latex, and the AcroScore system may require a learning curve that is simply too steep to engage many instructors. Although each of these programs is monetarily free, the time it takes to become proficient is a real barrier in a professional environment that may require teachers to use other course management software such as Moodle, WebWorks, and Blackboard. When confronted with this concern, Kaplan countered that using the tools he described may be the only tools an instructor needs because of their versatility and thus, they are well worth the investment in learning their uses.

By Cobb’s own admission his approach has challenges. A class environment where better students are expected to achieve at higher level than weaker students may not result in buy-in from all students, and may result in outright hostility. Such an approach hits directly a problem with higher education in general in that many students actively seek to achieve the lowest level of education return for the dollar as long as it is “certified” with a grade of an “A.” Some might argue that Cobb’s definition of fairness [“Fairness involves comparing students with each other.”] is too narrow and that a better definition involves comparison of each student to the stated goals and objectives of a course. We are sympathetic to this criticism, but we do feel that Cobb has highlighted a very intriguing way to challenge and stretch the most knowledgeable and gifted students in the class. In addressing this concern, Cobb notes that one has to engage the students in a dialogue directly and that in the upper level courses for which he has used this approach, buy-in has occurred. The issue of grade inflation is important as well. Such an approach where all students might earn A’s could result in repercussions for instructors in regard to promotion in rank.

We also note that the issue of praising students seems to be a fertile ground for much discussion among educators. Zieffler cited literature that claims praise inhibits learning, and yet Cobb and the AAHE principles advocate its use. We as instructors often wonder whether students reflect adequately (or at all) on the written feedback (positive or negative) we provide. It seems that personal teaching experience, anecdotal student stories, and the research literature form a “perfect storm” of confusion as to what may affect learning in the most positive way.

5.3 Possible Implementation Strategies

We advocate that instructors committed to assessment work together to review and consider implementing the ideas from Zieffler, Kaplan, and Cobb that resonate best for their needs. Certainly no one should be expected to attempt all these ideas in a short period of time, but to pick one or two ideas to integrate into an existing course. And instructors should not be expected to understand and adapt assessment theory on their own, but we strongly encourage a collaborative approach. In fact, it was a team approach that implemented the changes described by Zieffler.

After a systematic review of the course and the development of the assessments that accompany such a course, we urge such teams to make their efforts public. This could be done through the use of a course portfolio. Resources regarding portfolio development include works by Bernstein et al. (2006), Hutchings and Shulman (1998), the Peer Review of Teaching Project at the University of Nebraska, Lincoln (<http://www.courseportfolio.org>) and Seldin (2004). We believe that electronic publication of course or teaching portfolios could be extremely valuable to instructors of statistics. Possible venues for such publication include *The Journal of Statistics Education*, the CauseWeb website, or perhaps *Technology Innovations in Statistics Education*. Such a publication might differ from the typical article format, but it seems appropriate to us that these electronic formats expand to include other viable formats.

5.4 Final Thoughts

Zieffler, Kaplan, and Cobb have provided three excellent examples of systematic thinking concerning assessment, reminding us of the multiple dimensions of assessment, and challenging us to continually evaluate the goals and models of our assessment process. Each of the authors has shown us that assessment does have an impact on students’ experiences within a course. Although concerns remain about implementation, we commend the authors for their visions of the possibilities in using assessment in exciting new ways in their statistics courses. We encourage the statistics education community to continue to evaluate how assessment theory can be applied in our own courses.

[Received November 2008. Revised December 2010.]

REFERENCES

- American Statistical Association (2005), “Guidelines for Assessment and Instruction in Statistics Education: College Report,” Washington, DC, available at <http://www.amstat.org/education/gaise/>. [1]
- Angelo, T. A., and Cross, K. P. (1993), *Classroom Assessment Techniques: A Handbook for College Teachers* (2nd ed.), San Francisco: Jossey-Bass. [2]
- Arter, J., and McTighe, J. (2001), *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*, Thousand Oaks, CA: Cowin Press, Inc. [2]
- Astin, A. W., Banta, T. W., Cross, K. P., El-Khawas, E., Ewell, P. T., Hutchings, P. et al. (1992), “Nine Principles of Good Practice for Assessing Student Learning,” American Association of Higher Education Assessment Forum, available at <http://www.nwhealth.edu/ctl/asmnt/ninepgp.html>. [8]

- Atkin, J. M., Black, P., and Coffey, J. (2001), *Classroom Assessment and the National Science Education Standards*, Washington, DC: National Academies Press. [2]
- Bass, K. M., and Glaser, R. (2004), *Developing Assessments to Inform Teaching and Learning*, Center for the Study of Evaluation: National Center for Research on Evaluation, Standards, and Student Testing. [2]
- Bernstein, D., Burnett, A. N., Goodburn, A., and Savory, P. (2006), *Making Teaching and Learning Visible: Course Portfolios and the Peer Review of Teaching*, Boston: Anker. [9]
- Black, P., and Wiliam, D. (1998a), "Assessment and Classroom Learning," *Assessment in Education: Principles, Policy & Practice*, 5 (1), 7–71. [2]
- (1998b), "Inside the Black Box: Raising Standards Through Classroom Assessment," *Phi Delta Kappan*, 80, 139–148. [2]
- Bloom, B. S. (ed.) (1956), *Taxonomy of Educational Objectives: The Classification of Educational Goals: Handbook I, Cognitive Domain*, New York: Longmans, Green. [2]
- Butler, R. (1987), "Task-Involving and Ego-Involving Properties of Evaluation: Effects of Different Feedback Conditions on Motivational Perceptions, Interest, and Performance," *Journal of Educational Psychology*, 79 (4), 474–482. [3]
- Chance, B. L. (2000), "Experiences With Authentic Assessment Techniques in an Introductory Statistics Course," in *Teaching Statistics: Resources for Undergraduate Instructors*, ed. T. Moore, Washington DC: Mathematical Association of America, pp. 209–218. [1]
- (2004), *Proceedings of the ARTIST Roundtable Conference on Assessment in Statistics Education*, Lawrence University, available at <http://www.rossmanchance.com/artist/Proctoc.html>. [1]
- Corbett, A. T., and Anderson, J. R. (2001), "Locus of Feedback Control in Computer-Based Tutoring: Impact on Learning Rate, Achievement and Attitudes," in *Proceedings of ACM CHI 2001 Conference on Human Factors in Computing Systems*, New York: Association for Computing Machinery Press, pp. 245–252. [3]
- Downing, S. M. (2006), "Twelve Steps for Effective Test Development," in *Handbook of Test Development*, eds. S. M. Downing and T. M. Haladyna, Mahwah, NJ: Lawrence Erlbaum Associates, Inc., pp. 3–25. [2]
- Fedor, D. B. (1991), "Recipient Responses to Performance Feedback: A Proposed Model and Its Implications," *Research in Personnel and Human Resources Management*, 9, 73–120. [3]
- Gal, I., and Garfield, J. (eds.) (1997), *The Assessment Challenge in Statistics Education*, Amsterdam: IOS Press. [1]
- Gamoran, A., Porter, A. C., Smithson, J., and White, P. A. (1997), "Upgrading High School Mathematics Instruction: Improving Learning Opportunities for Low-Achieving, Low-Income Youth," *Educational Evaluation and Policy Analysis*, 19, 325–338. [2]
- Gardner, J. (ed.) (2006), *Assessment and Learning*, London: Sage Publications. [2]
- Garfield, J. (1994), "Beyond Testing and Grading: Using Assessment to Improve Student Learning," *Journal of Statistics Education*, 2 (1), available at <http://www.amstat.org/publications/jse/v2n1/garfield.html>. [1]
- Haladyna, T. M., Downing, S. M., and Rodriguez, M. C. (2002), "A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment," *Applied Measurement in Education*, 15 (3), 309–334. [2]
- Harlen, W. (2006), "On the Relationship Between Assessment for Formative and Summative Purposes," in *Assessment and Learning*, ed. J. Gardner, London: Sage Publications, pp. 103–117. [2]
- Hutchings, P., and Shulman, L. S. (eds.) (1998), *The Course Portfolio: How Faculty Can Examine Their Teaching to Advance Practice and Improve Student Learning*, Washington, DC: American Association for Higher Education. [9]
- Jordan, J. (2004), "The Use of Orally Recorded Exam Feedback as a Supplement to Written Comments," *Journal of Statistics Education*, 12 (1), available at <http://www.amstat.org/publications/jse/v12n1/jordan.html>. [2]
- Kluger, A. N., and DeNisi, A. (1996), "The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory," *Psychological Bulletin*, 119 (2), 254–284. [3]
- McCloskey, W., and Leary, M. R. (1985), "Differential Effects of Norm-Referenced and Self-Referenced Feedback on Performance Expectancies, Attribution, and Motivation," *Contemporary Educational Psychology*, 10, 275–284. [3]
- Mosteller, F. (1989), "The 'Muddiest Point in the Lecture' as a Feedback Device," *On Teaching and Learning. The Journal of the Harvard-Danforth Center*, 3, 10–21, available at <http://faculty.valenciac.edu/pnellis/units/IRCC/documents/mostellermuddypoint.pdf>. [2]
- Narciss, S., and Huth, K. (2004), "How To Design Informative Tutoring Feedback for Multi-Media Learning," in *Instructional Design for Multimedia Learning*, eds. H. M. Niegemann, D. Leutner, and R. Brunken, Munster, NY: Waxmann, pp. 181–195. [3]
- National Research Council (2001), *Knowing What Students Know: The Science and Design of Educational Assessment*, Washington, DC: National Academies Press. [2]
- Porter, A. C., Smithson, J., Blank, R., and Zeidner, T. (2007), "Alignment as a Teacher Variable," *Applied Measurement in Education*, 20 (1), 27–51. [2]
- Rodriguez, M. C. (2005), "Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research," *Educational Measurement: Issues and Practice*, 24 (2), 3–13. [2]
- Roseth, C. J., Garfield, J. B., and Ben-Zvi, D. (2008), "Collaboration in Learning and Teaching Statistics," *Journal of Statistics Education*, 16 (1), available at <http://www.amstat.org/publications/jse/v16n1/roseth.html>. [3]
- Roseth, C. J., Johnson, D. W., and Johnson, R. T. (2008), "Promoting Early Adolescents' Achievement and Peer Relationships," *Psychological Bulletin*, 134, 223–246. [3]
- Rowan, B. (1998), "The Task Characteristics of Teaching: Implications for the Organizational Design of Schools," in *Curriculum Leadership: Rethinking Schools for the 21st Century*, eds. R. Bernhardt, C. N. Hedley, G. Cattaro, and V. Svolopoulos, Cresskill, NJ: Hampton Press. [2]
- Sadler, D. R. (1989), "Formative Assessment and the Design of Instructional Systems," *Instructional Science*, 18, 119–144. [3]
- Schmeiser, C. B., and Welch, C. J. (2006), "Test Development," in *Educational Measurement* (4th ed.), ed. R. L. Brennan, Westport, CT: Praeger Publishers, pp. 623–646. [2]
- Seldin, P. (2004), *The Teaching Portfolio: A Practical Guide to Improved Performance and Promotion/Tenure Decisions* (3rd ed.), San Francisco: Jossey-Bass. [9]
- Shute, V. J. (2008), "Focus on Formative Feedback," *Review of Educational Research*, 78 (1), 153–189. [3]
- Stiggins, R. J. (2002), "Assessment Crisis: The Absence of Assessment for Learning," *Phi Delta Kappan*, 83 (10), 758–765. [2,3]
- Weldon, L., and Phillips, B. (eds.) (2007), *Assessing Student Learning in Statistics. Proceedings of the IASE/ISI-Satellite Conference*, Guimarães Cultural Centre Vila Flor Guimarães, Portugal, 19–21 August 2007, available at <http://www.stat.auckland.ac.nz/~iase/conferences.php?show=iasestat07>. [1]
- Western and Northern Canadian Protocol for Collaboration in Education (ed.) (2006), *Rethinking Classroom Assessment With Purpose in Mind*, Crown Right of the Government of Alberta, British Columbia, Manitoba, Northwest Territories, Nunavut, Saskatchewan, Yukon Territory: Western and Northern Canadian Protocol for Collaboration in Education. [3]
- Wiliam, D. (2007), "Keeping Learning on Track: Classroom Assessment and the Regulation of Learning," in *Second Handbook of Mathematics Teaching and Learning*, ed. F. K. Lester Jr., Greenwich, CT: Information Age Publishing, pp. 1053–1098. [3]
- Williams, S. E. (1997), "Teachers' Written Comments and Students' Responses: A Socially Constructed Interaction," paper presented at the Annual Meeting of the Conference on College Composition and Communication (48th, Phoenix, AZ, March 12–15, 1997), available at <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?acco=ED408589>. [3]