



# Multimedia Presentations in Educational Measurement and Statistics: Design Considerations and Instructional Approaches

Jeffrey C. Sklar  
California Polytechnic State University, San Luis Obispo

Rebecca Zwick  
University of California, Santa Barbara

*Journal of Statistics Education* Volume 17, Number 3 (2009),  
[www.amstat.org/publications/jse/v17n3/sklar.html](http://www.amstat.org/publications/jse/v17n3/sklar.html)

Copyright © 2009 by Jeffrey C. Sklar and Rebecca Zwick all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

---

**Key Words:** Test scores; Assessment; Pedagogy; Web-based.

## Abstract

Proper interpretation of standardized test scores is a crucial skill for K-12 teachers and school personnel; however, many do not have sufficient knowledge of measurement concepts to appropriately interpret and communicate test results. In a recent four-year project funded by the National Science Foundation, three web-based instructional presentations in educational measurement and statistics were developed and evaluated ([Zwick et al., 2008](#)). These modules were found to be particularly effective for pre-service K-12 teachers. The primary challenge of the project was to deliver the material in three short 25-minute web-based presentations. In this paper, we discuss the design principles, technical considerations, and specific instructional approaches implemented in the modules, invoking principles from cognitive psychology research. Based on evidence gathered from our project and previous research in teacher education and multimedia learning, we offer suggestions for presenting educational measurement and statistics concepts in a multimedia learning environment.

## 1. Introduction

Assessment literacy, loosely described as a teacher's competency in the principles and practices of testing and assessment, is a necessary skill for all teachers and school administrators. It has been noted that teachers spend up to one half of their time on assessment-related activities ([Stiggins, 1991](#)), and although interpreting test scores is only one component of assessment literacy, it is a vitally important one. As stated by [Impara et al. \(1991\)](#), "teachers cannot make valid use of scores if they do not understand the scores." In the wake of the No Child Left Behind Act of 2001, teachers have been under increased pressure to be able

to understand standardized test results and appropriately use those results to make classroom decisions.

The Instructional Tools in Educational Measurement and Statistics (ITEMS) project ([Zwick et al., 2008](#)) was designed to help K-12 teachers and school personnel acquire assessment literacy skills so that they would be better prepared to interpret standardized test results, use them to improve instruction, and communicate them to others. Three 25-minute web-based instructional presentations on various topics in educational measurement and statistics were developed and evaluated. The instructional videos were not designed to take the place of an entire course, but rather as a professional development activity for in-service teachers and school administrators, or as a supplement to relevant coursework taken by teacher education program (TEP) students. The challenge was to effectively and briefly communicate the material to a diverse audience, some of whom had very limited training in statistics and measurement. A primary finding was that the modules generally benefited teacher education program students more than in-service teachers and administrators.

In the first section of this article, we review the most relevant elements of the existing literature on assessment literacy. In the next section, we provide an overview of the ITEMS project. In Sections 3 and 4, we discuss module design considerations and describe the pedagogical approaches used to convey particular measurement and statistics concepts using animation, and in the final section, we provide a brief discussion of a potential new area of research. We conclude with some suggestions for presenting material in a computer-animated multimedia learning environment.

## 1.1 Assessment Literacy

Previous studies have found that teachers are deficient in the skills necessary to properly assess their students and to interpret standardized test results ([Pedulla, Abrams, Madaus, Russell, Ramos, & Miao, 2003](#); [Stiggins, 2002](#)). [Campbell & Evans \(2001\)](#) found that even after successfully completing a course in educational measurement, students enrolled in teacher education programs often failed to correctly apply measurement principles during their student teaching. In a recent investigation, researchers at the National Board on Educational Testing and Public Policy at the Lynch School of Education at Boston College surveyed a nationally representative sample of teachers to ascertain their attitudes about state-mandated testing programs. About a third of the respondents reported that their professional development in the area of standardized test interpretation was inadequate or very inadequate ([Pedulla et al., 2003](#)).

One explanation for the deficiency in assessment skills among in-service teachers is the general absence of a formal requirement in this area at the state and national level. In a recent study of teacher preparation in classroom assessment, [Stiggins and Herrick \(2007\)](#) found that no state required successful completion of a course in assessment for teacher certification. In addition, their study found that among the top ten teacher education programs (as ranked by *US News and World Report*), only one program (out of six that responded to repeated inquiries) required a stand-alone course in student assessment. They conclude that "...beginning teachers today may not be entering America's classrooms with the necessary knowledge and skills to use assessment FOR learning, much less be prepared to meet teacher standards in assessment" ([Stiggins & Herrick, 2007](#), p. 9).

Numerous opportunities have become available for teachers to learn about assessment outside of the traditional classroom setting, including books, DVD's, and workshops sponsored by institutes dedicated to assessment such as the Assessment Training Institute of Educational Testing Service. Some states have also implemented programs to assist pre-service and in-service teachers. [Lukin, Bandalos, Eckhout, & Mickelson \(2004\)](#) discuss several assessment training programs available to pre-service and in-service teachers in Nebraska, including the Pre-service Assessment Literacy Study Groups (PALS) and the In-service and Pre-service Assessment Literacy Study Groups (IPALS) programs, with results indicating increased knowledge and confidence in various areas of assessment literacy. Primary goals of these programs included developing classroom assessments and implementing the results for instructional

decision making, however; interpretation of standardized test scores was not a primary objective. [Wang, Wang & Huang \(2008\)](#) developed and evaluated the web-based system Practicing, Reflecting and Revisiting with Web-based Assessment and Test Analysis (P2R-WATA), designed to improve pre-service teachers' assessment literacy. Programs such as IPALS, PALS, and P2R-WATA are essentially full-length training courses. The instructional presentations from the ITEMS project also provide an opportunity for K-12 educators to improve their test score interpretation and assessment literacy skills, but are brief and self-contained.

## 2. Instructional Tools in Educational Measurement and Statistics

In each year of the ITEMS project, a web-based training module addressing different topics in educational measurement and statistics was developed and evaluated. All three modules are available for viewing on the ITEMS website:

<http://education.ucsb.edu/webdata/research/items/>

A brief summary of each module is provided below.

### 2.1. Module 1: What's the Score?

Two elementary school teachers, Stan and Norma, discuss introductory topics in statistics in the context of test scores, and interpretation of test results. Stan is somewhat tentative about tests at the beginning of the module; however, throughout the course of the presentation, learns about test score distributions and their properties (mean, median, mode, range, standard deviation), types of test scores (raw scores, percentiles, scaled scores, and grade-equivalents), and norm-referenced and criterion-referenced score interpretation.

Module 1 can be viewed on the web by visiting the link:

<http://education.ucsb.edu/webdata/research/items/modules/mod1/loader.html>

### 2.2. Module 2: What Test Scores Do and Don't Tell Us

In "What Test Score Do and Don't Tell Us," Stan from "What's the Score?" (who is now presumably better informed about standardized tests) meets with the parents of twins Edgar and Mandy, two of Stan's students. In this setting, Stan discusses the effect of measurement error on individual student test scores, test reliability, the effect of sample size on the precision of average scores for groups of students, and the definition and effect of test bias. Module 2 can be viewed at:

[http://education.ucsb.edu/webdata/research/items/modules/mod2/loader\\_high.html](http://education.ucsb.edu/webdata/research/items/modules/mod2/loader_high.html)

### 2.3 Module 3: What's the Difference?

In "What's the Difference?" a press conference is being held where the results from recently administered standardized tests are released. The superintendent of the school district, accompanied by Stan and Norma, is fielding questions from reporters about the importance of disaggregating test data for key student groups and the implications for score trend interpretation of shifts in the student population, the number of students assessed, and changes in tests and test forms. Module 3 can be viewed at:

[http://education.ucsb.edu/webdata/research/items/module\\_viewer/view/high.html](http://education.ucsb.edu/webdata/research/items/module_viewer/view/high.html) \*

### 2.4. Project Research Phases

A distinguishing feature of the ITEMS project was the incorporation of a research component in which the effectiveness of the modules was assessed. Each module was evaluated using a quiz containing questions tailored to the topics of the module. In the research phase for Module 1, the quiz (Quiz 1) contained 20

items. Quiz 2 contained 16 items, and Quiz 3 contained 14 items. Copies of the quizzes can be downloaded from the ITEMS website.

Individuals who agreed to participate in each research phase were initially asked to fill out an anonymous background survey in which demographic and employment information was collected. Participants were asked to indicate whether they were enrolled in a teacher education program. If not, they were asked to indicate their affiliation with K-12 education, that is, if they were teachers, principals, superintendents, or recently retired. (Educators who had been retired for more than five years were not eligible to participate.) Because few participants identified themselves as principals, superintendents, and other school personnel, these individuals were grouped with the teachers for most analyses. Hence, the two primary participant groups that were compared during the research phases were TEP students and school personnel.

Participants who agreed to take part in a research phase were randomly assigned to either take the quiz first (quiz-first group) and then view the presentation, or view the module first (module-first group) and then take the quiz. A total of 163 TEP students and 87 school personnel participated in the three research phases of the project. Table 1 provides the counts of TEP students and school personnel by module-first and quiz-first status for each research phase.

**Table 1:** *Number of participants by module research phase and group*

Research Phase	TEP Students		Non-TEP Students		Total
	Module-First	Quiz-First	Module-First	Quiz-First	
Module 1	33	35	19	26	113
Module 2	40	41	11	12	104
Module 3	8	6	10	9	33
Total	81	82	40	47	250

Average scores of the quiz-first and module-first groups were compared to determine the effectiveness of the modules. In general, school personnel outperformed TEP students, as measured by their average quiz score. However, the differences in average scores between the module-first and quiz-first groups, which provide an estimate of module effectiveness, were larger among TEP participants than among school personnel. Table 2 displays the average quiz scores for the quiz-first and module-first groups for all three quizzes. Among TEP participants, the effect sizes due to Modules 1, 2, and 3 were .35, .84, and .24 standard deviation units, respectively, while the corresponding effect sizes were .28, .10, and .20 among school personnel. These results suggest that TEP students were benefitting the most from the module presentations. Additional statistical analyses and a discussion of the results can be found in [Zwick et al. \(2008\)](#).

**Table 2:** Average score and standard deviations (in parentheses) by quiz and group

Research Phase	TEP Students			Non-TEP Students		
	Module-First	Quiz-First	All	Module-First	Quiz-First	All
Quiz 1: 20 items	13.1 (4.0)	11.7 (2.5)	12.4 (3.8)	13.4 (3.2)	12.5 (3.2)	12.9 (3.2)
Quiz 2: 16 items	12.6 (3.2)	9.5 (3.7)	11.0 (3.8)	12.7 (1.9)	12.5 (1.4)	12.6 (1.6)
Quiz 3: 14 items	6.5 (4.1)	5.5 (2.1)	6.1 (3.3)	11.2 (3.0)	10.4 (4.0)	10.8 (3.4)

Participants in each research phase also had opportunities to provide feedback on various aspects of the project including the module content, animation, and embedded questions, as well as whether the material would be useful to them. Comment boxes at the end of the module were available for participants to provide feedback regarding the quiz and presentation. In addition, if participants agreed, they were sent an additional evaluation survey to complete. Reports summarizing the evaluation results were composed by independent evaluators (see [Phillips, 2005](#); [Yeagley, 2006, 2007](#)).

### 3. Design Considerations

Our goal was to provide viewers a set of informative modules that were unlike existing videos that presented similar material using human actors and/or narrators. Animated presentations would allow flexibility in the choice and appearance of characters, and would allow us to illustrate and teach many concepts using creative graphics. In the following sections we describe the important design features of the modules, particularly the use of animated pedagogical agents, and the incorporation of multimedia design principles, navigational features, and embedded questions to periodically assess viewer learning.

#### 3.1. Animated Pedagogical Agents

Material was delivered through the use of animated pedagogical agents as an alternative to human actors. Animated agents are described as "computerized character[s] (either humanlike or otherwise) designed to facilitate learning" ([Craig, Gholson, & Driscoll, 2002](#)). Within multimedia learning environments, "animation can promote learner understanding when used in ways that are consistent with the cognitive theory of multimedia learning" ([Mayer & Moreno, 2002](#)), and have the potential to make the learning experience more engaging ([Gulz & Haake, 2006](#)).

Research into the benefits of animated pedagogical agents in a computerized environment has been a rapidly growing area. Multimedia environments with animated pedagogical agents introduce a social element into the teaching and tutoring process that can serve as a learning companion ([Kim & Baylor,](#)

[2006](#)). Research into the social cues of animated agents, also known as *social agency theory* ([Moreno, Mayer, Spires, & Lester, 2001](#)), has suggested that coordinating simple gestures and gazes into the agent's narration facilitates the learner's engagement with the lesson and the learning process ([Lusk & Atkinson, 2007](#); [Dunsworth & Atkinson, 2005](#); [Atkinson, 2002](#)). And implementation of well-designed animated pedagogical agents into computer environments "may provide learners with a sense of companionship and so make working in the computer-based environment relevant and/or meaningful" ([Kim & Baylor, 2006](#), p. 571).

[Clark & Choi \(2005\)](#) discuss three primary potential learning benefits linked to the use of agents: positive impact on learners' motivation, helping learners to focus on important elements of the material, and providing learners with context-specific strategies and advice. Additional empirical results have indicated that animation is superior to static images in terms of retention and transfer of information ([Mayer & Moreno, 2002](#); [Craig, Gholson, & Driscoll, 2002](#); [Moreno, Mayer, Spires & Lester, 2001](#)). [Baylor & Ryu \(2003\)](#) also investigated the roles of image and animation on TEP students' perceptions of the agent's characteristics, and found that it was important for the agent to be perceived as "instructor-like" and credible.

The primary animated characters used throughout the three ITEMS modules are two elementary school teachers, Stan and Norma. We wanted to maintain some consistency of the characters used throughout the modules, although it is not necessary to view the modules in consecutive order. A supporting cast of computer-animated students, parents, and a superintendent was also included throughout the modules. The ethnicities of the majority of the characters were intentionally ambiguous to provide a multicultural computer environment. All animations were produced using Macromedia Flash (now Adobe Flash) software.

### 3.2 Multimedia Design Principles

Multimedia learning refers to environments where learners are exposed to information in verbal (e.g., text or narration) and pictorial forms (e.g., static images or animation) ([Mayer & Moreno, 2002](#)). In designing our modules, we incorporated several multimedia principles discussed by [Mayer \(2001\)](#) and [Mayer & Moreno \(2002\)](#), which are briefly reviewed below.

**Multimedia principle.** According to this principle, "[s]tudents learn more deeply from animation and narration than from narration alone" ([Mayer & Moreno, 2002](#), p.93). Students can build better mental connections between words and pictures when both are presented, rather than having only one format presented. Research has indicated that "...human understanding occurs when learners are able to mentally integrate visual and verbal representations" ([Mayer, 2001](#), p. 5).

**Spatial and temporal contiguity principles.** The spatial contiguity principle states that students learn more deeply when on-screen text and animation are presented near each other ([Mayer & Moreno, 2002](#)). This principle applies primarily when the closed captioning option was used (to be discussed in Section 3.3). When this option is turned off, there is very little on-screen text since most of the material is presented using animation and narration (see the modality principle discussed below). The temporal contiguity principle states that "students learn more deeply when corresponding portions of the narration and animation are presented at the same time than when they are separated in time" ([Mayer & Moreno, 2002](#), p.95). Video and audio are synchronized throughout all presentations. Materials that incorporate these principles of temporal and spatial contiguity have been shown to enhance learning ([Mayer, 2001](#), pp. 81-112).

**Modality principle.** According to this principle, verbal information that accompanies graphical presentations should be presented in spoken form rather than as on-screen text. The learner's visual channel might become overloaded when words and pictures are presented simultaneously. The modality principle

([Mayer & Moreno, 2002](#); [Moreno & Mayer, 1999](#)) has been incorporated into multimedia design because it reduces what is known as the *split-attention* effect ([Chandler & Sweller, 1992](#)) which occurs when information from two or more sources needs to be processed simultaneously to derive meaning from subject material. (We made an exception to this principle in providing a closed-captioning option in Modules 2 and 3.)

**Coherence principle.** The coherence principle states that "students learn more deeply from animation and narration when extraneous words, sounds, and video are excluded" from the presentation ([Mayer & Moreno, 2002](#), p. 95). Irrelevant material has been shown to interfere with the learning process. Hence, efforts were made to remove extraneous words, pictures, and sounds from the modules, although we did occasionally introduce a few entertaining remarks.

**Personalization principle.** The personalization principle states that learning is facilitated when narration is conducted in a conversational style rather than a formal style, and students will be more motivated when they are personally involved in the conversation ([Mayer & Moreno, 2002](#)). Hence, a conversational style (incorporating pronouns such as "I," "you," and "we") was used in each module.

### 3.3. Navigational Features

Over the course of developing the three modules, we added specific navigational features that allowed the viewers some control over the pace of the presentations. The original version of Module 1 (no longer available for viewing) had the fewest options for viewers to control the scenes. Viewers could not pause or skip to different parts of the presentation, and, hence, could not easily access specific parts of the module to review material. Originally, for evaluation purposes, we wanted all participants to have an "equal" viewing experience, controlling as many extraneous factors as possible that might confound any effects on assessment quiz score attributed to the module.

The feedback that we received from both participants and project advisory committee members indicated that navigational revisions were in order. As one participant stated: "In order to be an effective teaching tool I do think that one would want to have the opportunity to revisit the presentation ..." Another viewer commented that "If there was a pause button on the presentation it may help teachers who experience a distraction during the presentation not to miss anything. Another solution may be to allow the teachers [to] control the presentation [by] rewinding a section in order to better understand what was being said by reviewing the information." Keeping these comments, as well as others, in mind, we revised many navigational aspects of Module 1 and also added "dividers" between the main content scenes of the module. Viewers can now switch to different scenes; however, it is still not possible to pause the presentation or perform short incremental forward or backward skips.

The improvements in the navigational features were also implemented in Modules 2 and 3. In Module 2, a "Main Menu" screen is available that allows the viewer to instantly skip to one of six separate scenes (an introduction, four content chapters, and a conclusion), and within each scene, the viewer has the options of pausing the presentation, skip-rewinding, skip-fastforwarding, or returning to the Main Menu. A closed-caption option was added as well, as an aid to viewers who are hard of hearing or do not have satisfactory audio equipment.

### 3.4. Embedded Questions

An additional pedagogical improvement that was implemented in Modules 2 and 3 was to include "embedded" questions at the end of each of the content scenes. As one viewer of Module 1 remarked, "it would be helpful if after each section there was a mini-assessment to see how well I absorbed the information. What would be great would be a semi-interactive format where after each section, there was a mini-assessment and the information that was not correct would be reviewed again."

Beginning with Module 2, we included a short quiz question following each content scene. A question with multiple choice answers is displayed on screen, and the viewer is given the opportunity to select an answer. Regardless of the answer chosen by the viewer, the narrator provides the correct answer and a brief explanation shortly after the question is displayed. Viewers who still do not understand the concept are advised to watch the relevant section again. This aspect added an interactive element to the module which has been shown to be effective for retaining information in studies on interaction with pedagogical agents ([Moreno et al, 2001](#)).

## 4. Pedagogical Approaches

In this section we discuss the instructional and pedagogical approaches that were used in the modules and provide some specific examples. We used both static graphs and animation to illustrate mathematical procedures and statistical concepts. We created realistic test score reports to illustrate measurement principles. Finally, we used analogies to help viewers connect their prior knowledge to new concepts.

### 4.1. Excluding Formulas

In keeping with our goal of maintaining a conversational style, we made the decision to exclude formulas from the instructional modules. (The relevant formulas are given in supplementary materials available at <http://education.ucsb.edu/webdata/research/items/pages/modules.html>) In their place, we substituted traditional formulas with animation and graphs that would "mimic" mathematical operations. Our philosophy of presentation is much the same as that of the statistics textbook by authors [Freedman, Pisani, Purves and Adhikari \(1991\)](#), p. xiii), who state that "Mathematical notation only seems to confuse things for most people, so we [explain statistics] with words, charts, and tables—and hardly any x's or y's ... What [people] really need is a sympathetic friend who will explain the ideas and draw the pictures behind the equations. We are trying to be that friend..." Although some of the statistical concepts were introductory, such as mean, median, and standard deviation, others might be regarded as intermediate, such as reliability, standard error of measurement, and Simpson's Paradox.

An example of an intermediate topic illustrated in Module 2 is the effect of an individual test score on the average score for an entire class or school. (See [http://education.ucsb.edu/webdata/research/items/modules/mod2/loader\\_high.html](http://education.ucsb.edu/webdata/research/items/modules/mod2/loader_high.html), Part 3, "How Does the Number of Students Tested Affect Imprecision in Average Test Scores?") In this sequence of images, viewers observe that in a small class, one student with a high test score can have a large impact on the average score, while removing a single student's score from a large class has little impact on the average.

Simpson's Paradox, sometimes called the amalgamation paradox, is illustrated in Module 3 (though neither of those terms is used). This paradox is a phenomenon that occurs when the direction of an association between two variables is reversed when a third variable is controlled ([Utts & Heckard, 2004](#) for examples). An example of the paradox is provided in the context of test scores. The proficiency rate increases at a particular school from 30% to 35% from one year to the next for students in an economically disadvantaged group and from 78% to 80% in the non-disadvantaged group; however, the *overall* proficiency rate decreased from 73% to 71% for all students combined. (See [http://education.ucsb.edu/webdata/research/items/module\\_viewer/view/high.html](http://education.ucsb.edu/webdata/research/items/module_viewer/view/high.html), Part 2, "How Do School Population Changes Affect Test Score Trends?")

### 4.2 Graphs

Simple graphical displays were useful for conveying statistical concepts, including variability, percentage of correct responses, percentiles, and percentile ranks. Variability is an important fundamental concept that is often overlooked by teachers in the context of test scores. In a study of mathematics teachers' notions of

variability ([Makar and Confrey, 2002](#)), teachers used *Fathom* (Finzer, 2001) statistical software to investigate variability in Texas high-stakes test data. When comparing sets of test score data, teachers tended to describe the centers of the groups of data, but generally ignored any discussion of variability within and between the sets of data.

In our discussion of variability in test score distributions in Module 1, several illustrations were provided. For example, we displayed two score distributions with the same center but different spreads. (See <http://education.ucsb.edu/webdata/research/items/modules/mod1/loader.html>, Part 2, "How Can I Describe the Spread of the Scores?") Our techniques for communicating variability were in line with those discussed in the statistics education literature. Specifically, [Garfield and Ben-Zvi \(2005\)](#) note that "When making comparisons of two or more data sets, examining their graphs on the same scale allows us to compare the variability and speculate about why there are differences."

In Module 1, graphs were also useful for illustrating percentiles, percentile ranks, and highlighting differences between percentiles and percentage correct, two topics that are commonly confused. (See <http://education.ucsb.edu/webdata/research/items/modules/mod1/loader.html>, Part 5 "What's a Norm-Referenced Score Interpretation?")

### 4.3 Analogies

To facilitate the presentation of new and potentially complex topics in a restricted time frame, familiar situations (i.e., analogies) were used to link prior knowledge concepts to new material. Several empirical investigations have suggested that the use of analogies in instruction aids student learning (e.g., [Bulgren, Deshler, Schumaker & Lenz, 2000](#); [Simons, 1984](#)).

To introduce the concept of variability in Module 1, we used a "spread" analogy by presenting an illustration of butter being spread on toast to represent the spread of test scores. This introduction is less conventional than what would typically be given in a lesson on variability, and its primary purpose was to link a familiar concept to a new one. This scene was subsequently followed up with graphs of distributions with different spreads. (See <http://education.ucsb.edu/webdata/research/items/modules/mod1/loader.html>, Part 2 "How Can I Describe the Spread of the Scores?")

Another example of an illustrative analogy that was used to introduce a statistical topic is provided in Module 1. To convey the concept of a distribution of test scores, Norma throws stacks of test score reports corresponding to individual scores next to each other resulting in an image of a histogram. Teachers are likely to be familiar with stacks of score reports and then quickly make the connection to the concept of a distribution as a method of summarizing a set of data. (See <http://education.ucsb.edu/webdata/research/items/modules/mod1/loader.html>, watch Part 1 "How Can I Get an Overview of the Test Scores of My Class?")

Analogies were used to introduce topics in educational measurement as well. In Module 2, an analogy was used to convey measurement error. Two side-by-side scales are displayed each weighing a candy bar, but showing two different weight readings. The scene illustrates that, because of imprecision in measuring capabilities, different results may be obtained on different measurement occasions. This is similar to the imprecision involved in using educational tests to measure student skills. Although viewers may be unfamiliar with the term "measurement error," they are probably familiar with weighing themselves on two different scales and observing different weights. (See [http://education.ucsb.edu/webdata/research/items/modules/mod2/loader\\_high.html](http://education.ucsb.edu/webdata/research/items/modules/mod2/loader_high.html), Part 1, "Why aren't test scores perfect measures of academic ability?")

## 4.4 Realistic Test Score Reports

One final important aspect of the presentations that we believed would aid understanding of test-interpretation was the use of realistic test score reports as a basis for explaining concepts and terminology. This is consistent with a "case-based" approach to learning emphasized in [Lundeberg, Levin, and Harrington \(1999\)](#). Two fictitious tests, the State Test of Academic Knowledge (STAK), a criterion-referenced test, and the National Education Achievement Test (NEAT), a norm-referenced test were used to illustrate many concepts throughout the module. Figure 1 displays STAK results for Steve, 6<sup>th</sup>-grader. This score report is used for discussing several different topics, including raw and scaled scores, and proficiency cut-scores.

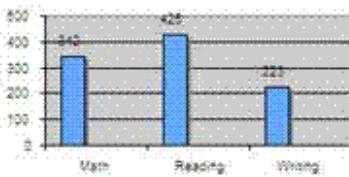
### PART 1: RESULTS FOR THE STAK: STATE TEST OF ACADEMIC KNOWLEDGE

Note: The STAK is administered to all students in grades 3-8 in the state. Scores range from 0 to 500 in each grade and subject area. The state has determined that a score of 300 is needed in order to consider a student proficient. The state's achievement levels are defined as follows:

0-149	BELOW BASIC
150-299	BASIC
300-449	PROFICIENT
450-500	ADVANCED

### REPORT FOR: STEVE (GRADE 6)

Test	Number of Questions	Raw Score (Number Correct)	Percent Correct	Scaled Score
Math	63	50	79%	375
Reading	40	36	90%	440
Writing	30	16	53%	240



**Figure 1:** STAK score report for Steve

## 5. Discussion

The goal of the ITEMS Project was to create short animated presentations that would assist pre-service and in-service teachers, as well as school administrators with interpreting standardized test results. A major challenge associated with developing the videos was the need to present complex topics in a constrained time period. Topics that would normally be covered in one or two hour-long class lectures (e.g., reliability) were covered in two to three minutes. In this paper, we have examined several design features and instructional approaches used to present measurement and statistics topics in a multimedia based environment.

One of the most important design considerations for the modules was to use animated pedagogical agents to deliver the material. Although our decision to do so was grounded in research, it is not entirely clear in retrospect that animation was the optimal way to deliver the material. Animation provided a method by which we could easily demonstrate complex topics without using formal lectures; however, some viewers found the animation to be distracting. As one viewer of Module 1 noted: "I found that I was more

entertained by the animation of the woman hurling papers across the room into perfect stacks, wishing that I could have such fantastic organizational skills that I forgot to pay close attention to what the video was saying." In addition, a small number of viewers commented that they found the use of animation to be condescending, with one viewer of Module 3 stating "I felt it was addressing children and not adults."

Unfortunately, we are limited in the conclusions we can draw about the effectiveness of the implemented multimedia principles. To assess the effectiveness of each principle would have required developing and evaluating several different versions of the modules, which were not original goals of the project. However, our finding that the effect sizes for TEP students were larger than those for teachers and administrators is consistent with earlier research indicating that the modality and contiguity effects are stronger for learners with low content knowledge (Moreno & Mayer, 1999). In other words, implementation of the multimedia principles does appear to benefit individuals whose familiarity with the presented topics is limited.

It was also not possible to empirically address whether the implemented navigational components improved or enhanced learning; however, the viewer feedback concerning these options was generally positive. Regarding the closed captioning, a Module 3 viewer (who incidentally did enjoy the animation) commented: "I liked the animations and also how there were captions below the screen." Having the option to read the dialogue may help individuals if they were not able to clearly hear the terms being discussed. The short assessments at the end of each content section also appeared to benefit the viewers, with a Module 3 viewer stating that "the video was informational and the questions throughout helped clarify the information that was being presented." During the Module 3 research phase, we also surveyed participants to determine their opinion about the embedded questions. Of the 32 participants who responded, 30 found the embedded questions "somewhat helpful" or "very helpful," while only 2 viewers indicated they were "neither annoying nor helpful." None considered them "distracting or annoying."

## 6. Conclusion

We consider this paper to be a starting platform for those who are interested in conducting studies on current and potential pedagogical approaches for teaching topics in educational measurement and statistics in a multimedia environment. As more opportunities come online for teachers and other school personnel to develop and improve their assessment literacy skills, more empirical research is needed to understand the best methods for teaching TEP students and in-service teachers measurement and assessment skills. The following is a list of recommendations for designing animated presentations in measurement and statistics and instructional approaches for teaching the material:

- Presentations should implement multimedia design principles.
- Topics should be presented in clearly partitioned scenes rather than one single continuous presentation.
- Navigational features should be incorporated: viewers should have some control over the pace of the presentation so that they can re-examine particular scenes or pause the presentation to absorb material.
- Embedded questions should periodically be provided during the presentation to assess viewer's understanding of the content.
- Complex mathematical equations and computations should be avoided. Instead, dynamic images or static graphs that can mimic mathematical operations should be used.
- Analogies should be used to invoke prior knowledge.
- Test score reports that contain information similar to that presented in actual score reports should be used as illustrations.

In conclusion, we have raised some issues pertaining to principles, strategies, and pedagogical approaches for teaching concepts in educational measurement and statistics in a computer-based environment. Whether implementation of some or all of these approaches enhances learning is not yet clear. A primary goal of our

discussion is to stimulate ideas for future experimental designs into the effectiveness of instructional approaches to teaching educational measurement in a computer based environment. Future research should focus on empirical investigations of these design features and instructional approaches to examine their effectiveness.

---

## Acknowledgments

We appreciate the support of the National Science Foundation (#0352519). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

---

## References

Atkinson, R.K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, 94(2), 416-427.

Baylor, A.L. & Ryu, J. (2003). The effects of image and animation in enhancing pedagogical agent persona. *Journal of Educational Computing Research*, 28(4), 373-394.

Bulgren, J. A., Deshler, D. D., Schumaker, J.B. & Lenz, B. K. (2000). The use and effectiveness of analogical instruction in diverse secondary content classrooms. *Journal of Educational Psychology*, 92(3), 426-441.

Campbell, C. & Evans, J.A. (2001). Investigation of preservice teachers' classroom assessment practices during student teaching. *The Journal of Educational Research*, 93(6), 350-355.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics, *Technology Innovations in Statistics Education*, 1(1).  
<http://repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art2>

Chandler, P. & Sweller, J. (1992). The split attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62, 233-246.

Clark, R.E. & Choi, S. (2005). Five design principles for experiments on the effects of animated pedagogical agents. *Journal of Educational Computing Research*, 32(3), 209-225.

Craig, S.D., Ghobson, B., & Driscoll, D.M. (2002). Animated pedagogical agents in multimedia environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, 94(2), 428-434.

Dunsworth, Q. & Atkinson, R. K. (2007). Fostering multimedia learning of science: Exploring the role of an animated agent's image. *Computers and Education*. 49(3), 677-690.

Freedman, D. Pisani, R., Purves, R., & Adhikari, A. (1991). *Statistics* (2nd ed). New York: W.W. Norton.

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, 4(1), 92-99.

Gulz, A. & Haake, M. (2006). Design of animated pedagogical agents- A look at their look. *International Journal of Human-Computer Studies*, 64(4), 322-339.

Impara, J.C., Divine, K.P., Bruce, F.A., Liverman, M.R., & Gay, A. (1991). Does interpretive test score information help teachers? *Educational Measurement: Issues and Practice*, 10(4), 16-18.

Kim, Y. & Baylor, A.L. (2006). A social-cognitive framework for pedagogical agents as learning companions. *Educational Technology Research and Development*, 54(6), 569-596.

Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, 23(2), 26-32.

Lundeberg, M. A., Levin, B. B., & Harrington, H. L. (Eds.). (1999). *Who learns what from cases and how? The research base for teaching and learning with cases*. Mahwah, NJ: Erlbaum.

Lusk, M.M. & Atkinson, R.K. (2007). Animated pedagogical agents: Does their degree of embodiment impact learning from static or animated worked examples. *Applied Cognitive Psychology*, 21(6), 747-764.

Makar, K. & Confrey, J. (July, 2002). Comparing two distributions: Investigating secondary teachers' statistical thinking. Presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa. Available at [http://www.stat.auckland.ac.nz/~iase/publications/1/10\\_18\\_ma.pdf](http://www.stat.auckland.ac.nz/~iase/publications/1/10_18_ma.pdf)

Mayer, R. E. (2001). *Multimedia learning*. Cambridge, UK: Cambridge University Press.

Mayer, R.E. & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review*, 14(1), 87-99.

Moreno, R. & Flowerday, T. (2006). Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology*, 31(2), 186-207.

Moreno, R. & Mayer, R.E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91(2), 358-368.

Moreno, R., Mayer, R., Spires, H., & Lester, J. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 19, 177-213.

Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Phillips, L. (2005). *Formative evaluation report* (No. 2): UCSB Gevirtz Graduate School of Education.

Popham, W.J. (2006). Needed: A dose of accountability. *Educational Leadership*, 63(6), 84-85.

Simons, P. R. (1984). Instructing with analogies. *Journal of Educational Psychology*. 76(3), 513-527.

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

Stiggins, R. (March 13, 2002). Assessment for learning. *Education Week*, vol. 21, no. 26, pp. 30, 32-33.

Stiggins, R. & Herrick, M. (2007). A status report on teacher preparation in classroom assessment. Unpublished manuscript.

Utts, J. M. & Heckard, R. F. (2004). Mind on Statistics (2nd Ed). Belmont, CA: Brooks/Cole.

Wang, T-H, Wang, K-H & Huang, S-C. (2008). Designing a Web-based assessment environment for improving pre-service teacher assessment literacy. *Computers and Education*, 51(1), 448-462.

Yeagley, P. (2006). Project evaluation of year 2: Instructional tools in educational measurement and statistics (ITEMS) for school personnel. Unpublished report.

Yeagley, P. (2007). Project evaluation of year 3: Instructional tools in educational measurement and statistics (ITEMS) for school personnel. Unpublished report.

Zwick, R., Sklar, J., Wakefield, G., Hamilton, C., Norman, A., & Folsom, D. (2008). Instructional tools in educational measurement and statistics (ITEMS). *Educational Measurement: Issues and Practice*, 27(2), 14-27. [Erratum for Zwick et al., *Educational Measurement: Issues and Practice*, 27(2), 14-27 in *Educational Measurement: Issues and Practice*, 27(4)].

---

Jeffrey C. Sklar  
Statistics Department  
California Polytechnic State University, San Luis Obispo  
1 Grand Ave.  
San Luis Obispo, CA 93407  
Email: [jsklar@calpoly.edu](mailto:jsklar@calpoly.edu)  
Telephone: 805-756-6353

Rebecca Zwick  
Gevirtz Graduate School of Education  
University of California, Santa Barbara  
Santa Barbara, CA 93106  
Email: [rwick@education.ucsb.edu](mailto:rwick@education.ucsb.edu)  
Telephone: 805-893-7762

---