Paper presented at the ARTIST Roundtable Conference on Assessment in Statistics
held at Lawrence University, August 1-4, 2004

Assessment of Using Technology for Teaching Statistics

Presentation for ARTIST Conference
August 1-4, 2004

Ginger Holmes Rowell, Ph. D.
Middle Tennessee State University

Abstract

The list of technologies used in statistics classes today has been expanded from those used in a classroom twenty-five years ago. Today's instructional technologies include Java applets for simulating and visualizing concepts in probability and statistics, free Internet-based analysis tools for calculating test statistics and p-values, multimedia textbooks which show video clips of real world situations that use statistics, and easy access to realistic data sets which can be downloaded and used in one of the many statistical application software packages. With so many options of technologies from which to select, a statistics teacher needs to have a means of assessing the effectiveness of these instructional technologies. This paper presents several general strategies associated with assessing instructional technologies. Specific examples are drawn from concepts and technologies used in teaching statistics. Current efforts of organizations to assist statistics teachers with the assessment of the wide array of materials available on the Internet are also presented.

Introduction

The approaches used for teaching statistics have changed in the past two decades with an increased focus on understanding concepts, using active learning strategies, and using real world data. (Gordon, 1992; Moore, 2000; Rossman, 1996-1999; Snell, 1999) Especially in recent years, advances in technology and in particular computer-related technologies have opened the doors for opportunities for curricular changes in both the methods and the materials used to teach statistics. Statistical software packages and graphing calculators that can conduct statistical hypothesis tests have been a part of the shifting emphasis from memorizing formulas to writing correct interpretations of the computer or calculator output. Increased speed of and improved access to the Internet have allowed for additional innovations which affect teaching: Internet-based applets are designed to enable students to quickly simulate random processes, Internet-based animations are developed to provide visualizations of concepts in probability and statistics, Internet-based analysis tools are freely available for calculating test statistics and p-values, and up-to-date real-world data sets are accessible from numerous Internet sources.

The issue of using technology as a tool for teaching statistics received international attention in 1996 as statistics education leaders discussed the "Role of Technology" at the Roundtable Conference sponsored by the International Association for Statistical Education (IASE). In the preface to the conference proceedings, Garfield, the Program Chair, highlights the need for valid, informative assessment methods for measuring "changes in students' understanding as they interact with technology." (1996, p. 2) She also indicates an awareness of the challenges associated with conducting research on the role of technology and reports that much of the research at that time used individual interviews or other procedures that do not transfer well to large classes. The statistical education leaders at this conference believed that future research should provide a "deeper understanding of statistics learning and thinking processes in technological settings." (Garfield, 1996, p. 2) Over time, such research would provide best practices associated with using technology as a tool for teaching and learning probability and statistics. Developers of instructional technologies for these disciplines could then use these best practices to guide their development process, resulting in improved methods and materials for teaching probability and statistics in the future.

This paper describes several strategies for assessing instructional technologies in a general setting. Next, examples related to concepts and technologies used for teaching statistics are presented. Though there are many types of instructional technologies, teaching materials that can be disseminated on the Internet are of particular interest in today's educational community. As such, current efforts of organizations to assist statistics teachers with assessing this wide array of materials available on the Internet are also discussed.

Some General Approaches for Assessing Instructional Technology

Organizations, consortium, and even companies have been established with the sole purpose of evaluating and assessing the use of instructional technology. By bringing together a large cohort of experts, these groups can conduct large-scale research projects in an effort to help establish best practices in the area of instructional technologies. Three such groups are described along with some different aspects associated with their approaches to assessing instructional technologies.

The *Teaching, Learning, and Technology Group* (TLT Group) is an example of a non-profit corporation which offers tools, information, and training to help improve teaching and learning through "thoughtful use" of information technology and to help "accelerate educational improvement while easing the stresses of institutional change." (TLT, n.d., p. 1) The TLT Group has provided their services to over 500 educational institutions and corporations worldwide. The TLT Group's *Flashlight Program* focuses on assessment and earns its name because it uses assessment to help see how technology is being used to improve education. Subscribing institutions are given access to over 20 different resources such as tool kits, study packages, rubrics, and tutorials related to assessment. *Flashlight* also provides consultants, training, and external evaluators. Steven W. Gilbert, the president and founder of the TLT Group, has an audio enhanced PowerPoint presentation on the group's website describing three types of assessment approaches for instructional technologies. (Gilbert, n.d.)

In this presentation, he explains that assessment of technology has progressed and developed more fully over time. The most simplistic assessment approach he describes is a *Monad Approach*, which deals only with the technology aspect and leaves out the evaluation process. Monadic thinkers select technology for technologies sake. The *Dyadic Approach* combines an emphasis on the technology being used with the outcome that is to be achieved. Evaluation occurs by assessing if this outcome is achieved. In cases when only one specific use of the technology is expected, then the dyadic approach may be sufficient. However, in many instances related to teaching, technology will have multiple uses or activities associated with it. Gilbert and his colleagues at TLT prefer the *Triad Approach* for assessing instructional technologies. This three-part approach is similar to the dyadic approach but adds the component of focusing on the specific activity or activities that use the technology to produce an outcome. The activity is what the user does with the technology, and the result of the activity is the outcome. The corresponding assessment is designed to determine if the use of technology was effective in producing the desired learning outcome when applied to a specific learning task. Gilbert believes that considering these three components (technology, activity, and outcome) and how they work together is important to help predict whether or not the technology fosters learning. (Gilbert, n.d.)

A second group striving to improve instructional technologies through assessment is the *Network for the Evaluation of Education and Training Technologies* (EvNet). EvNet is a national multi-discipline, multi-sector network (including universities, community colleges, schools, private, public, and non-profit organizations) committed to improving instructional technologies used in the Canadian education system through research focused on assessment and evaluation. EvNet provides customized research and evaluation services not only to schools but also to other training organizations. EvNet has developed an Evaluation Toolkit, which includes paradigms, questionnaires, bibliographies and rating criteria. EvNet's director, Carl Cuneo, presents twenty standards for best and worst practices for assessing instructional technologies. These standards use word pairs such as *inclusively/exclusivity*, *affordable/costly*, *learner active/learner passive*, etc. to convey the best/worst practices respectively. (Cuneo, 2004) Table 1 (on the following page) is a reorganization of some descriptions and labels EvNet uses to indicate best and worst practices for evaluating instructional technologies including courseware. (EvNet, p. 2-3) These labels can be used to assist educators in the selection of appropriate technology tools for their own course/program needs. In the format below, the labels act as a relatively easy-to-use checklist, which provides general guidance for instructional technology developers.

Table 1. EvNet's Best and Worst Practice Labels (reordered, EvNet, p. 2-3)

Best Practice Labels	Worst Practice Labels
friendly and intimate; people centered (warmware)	• 'coldware', user unfriendly • distancing (technology creates barriers between people who prefer distant e-mail to personal contacts)
inclusivity, equity of access (gender, race, income, region)	• blocks access, or differentiates access by social characteristics (income, employment status, age, region, ethnicity, gender) • expensive and costly
• enhances control by the user • learner driven • enhances user control over time; self paced, not instructor paced or technology paced (any time, any place learning)	• loss of control • loss of privacy
• knowledge building, creative, artistic, constructive • extending the senses	• 'high powered drill and kill' programs • skill destroying • electronic gaming effects (violence; sexism) • destructive of artistic talent; loss of art forms and culture
• creates opportunities for meaningful interaction	• unidirectional (teacher to student sponge lecturing); no interactivity (common in multimedia presentations) • destruction of community and collaborative learning
multidimensional (text, sound, graphics, video and animation in a balanced symphony that is not overpowering)	• unidimensional (text or graphics or sound or animation or video) • 'old wine in new bottles': repackaging old material with new graphics and sound bites on the Internet
eliminates routine tasks, allowing more time for higher order thinking and learning	• random, aimless browsing of the Internet; time waster • speed for speed's sake; pressure to respond now to e-mail
enabling (use of technology to overcome handicaps)	
democratizes and deregulates the educational experience	
	lack of personal support for finding solutions to computer glitches

A third organization dedicated to providing research-based resources to educators and policy makers is *The North Central Regional Educational Laboratory* (NCREL), a subsidiary of Learning Point Associates. NCREL, a member of the Regional Educational Laboratory Network, focuses its efforts in the states of Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin. Gilbert Valdez, Director of the Teaching, Learning, and Curriculum Center at the NCREL, provides a guide of the basic steps for evaluation of using technology:

- Identify basic issues/problems that need attention
- Develop revision strategies and priorities
- Determine what data will be necessary to obtain in order to make informed decisions
- Determine what data is useful for evaluation that already exists
- Determine what new data will need to be collected
- Identify evaluation design
- Determine what tools and data acquisition strategies will be used
- Determine timelines and specific people responsible for evaluation
- Conduct evaluations
- Analyze data and develop conclusions
- Report data in ways appropriate to various audiences. (Valdez, n.d., p. 2)

Valdez points out that the evaluation of educational technologies should provide information about outcomes produced by various applications of technologies, such as changes in student achievement, attitudes, engagement, and self-efficacy. Valdez suggests these evaluations should include a mix of quantitative and qualitative methods including case studies to analyze the context in which the technology is applied. (n.d., p. 3)

The international statistics education community also brought out the value of using multiple methodologies for the assessment of teaching/learning using technology. In his discussion of the "Empirical Research on Technology and Teaching Statistics" at the 1996 IASE Roundtable Conference, Shaughnessy advises those researching using technology to teach statistics to use several methodologies simultaneously. He suggests that this approach is particularly beneficial for investigating student thinking and reasoning. Using a sample of student interviews is a suggestion to supplement written responses collected from the students in order to benefit from having detailed information from student responses and access to large samples of data.

Examples Related to Assessment of Technology for Teaching Statistics

The general methods for assessing instructional technologies can be used to make determinations about the effectiveness of the use of technologies for teaching statistics. This section begins with a very specific example of an application of the Triad Approach suggested by Gilbert, which also serves as a reminder that "technology" can mean more than just high-tech methods. Second, an example of defining best features associated with specific software for developing understanding of a statistical concept is presented.

First, we will revisit the Triad Approach to assessing technology in which the user begins with a desired outcome, then selects the learning activity believed to help achieve the outcome, and finally selects the technology appropriate for the task. An example related to statistics begins with a teacher selecting a desired outcome for his/her students.

Assume the desired outcome is that the students will be able to recognize convergence of the empirical distribution to the theoretical distribution for the binomial distribution. The teacher may also want the students to be able to visualize the shape of the binomial distribution for different values of the parameters *n* and *p* (the number of trials and the probability of success in a binomial experiment, respectively). Next, the teacher will select an activity believed to provide the desired outcome. In this case, perhaps the activity selected is to simulate a specific binomial experiment 10, 100, and then 1000 times and to vary *n* and *p.* Third, a technology is selected that is appropriate for the activity. Thus, for 10 simulations, the teacher may devise a low-tech simulation in which the students simulate the process by selecting items from a bag or some other tactile simulation. However, for simulating a process 1000 times a computer-aided simulation would be selected. The teacher may choose an Internet applet such as a Binominal Experiment from "The Virtual Probability and Statistics Laboratory" (Siegrist, 1997-2004) and have the students use this applet to complete a series of tasks defined by the teacher, or the teacher may have students program a simulation using some type of statistical software. The teacher uses his/her own expertise about the subject and knowledge about the abilities of their students to select the appropriate technology based on the activity and desired outcome. The teacher would assess the students using perhaps even some traditional methods (*e.g*. quiz, test, homework) to see if the learning objectives were met. The teacher may also question the students to gain extra insight into the use of the technologies (high-tech and/or low -tech) for achieving the desired learning outcomes. This questioning could be a survey, an interview, or even a simple two-question quiz related to the technology usage. As suggested by Valdez, Shaughnessy, and Garfield multiple methods of assessment can provide additional insight into student understanding of what is happening using technology.

At the 1996 ISAE Roundtable Conference on the Role of Technology and revisited at the 2002 Joint Statistical Meetings panel discussion[1], delMas described his efforts to develop and assess statistical software. delMas is very thorough and deliberate in his development of his evaluation strategy. His objective is that students be able to explore simulations of sampling distributions to learn and/or discover Central Limit Theorem concepts. delMas carefully defined the activities that he wanted the students to do and then created his own computer software, called *Sampling Distributions*, to meet his objectives. To evaluate and assess this software, delMas combines existing research models, Nickerson and Holland et al. "Inductive Reasoning Model," and comes up with the following list of software features to promote understanding:

1. The program should start where the student is, not at the level of the instructor. The program should accommodate students' conceptual development in a domain, common misconceptions that occur, and the knowledge that students typically bring to the classroom.
2. The program should promote learning as a constructive process in which the task is to provide guidance that facilitates exploration and discovery by
    (a) providing ways for the student to make and test predictions,
    (b) promoting frequent generation and testing of predictions, and
    (c) providing clear feedback on the outcomes of predictions.
3. The program should use models and representations that are familiar to the novice. Novices tend to generate concrete representations based on familiar concepts or

[1] "Our Toolbox is Overflowing: Toward a Framework for Mapping Technologies and Topics in Introductory Statistics."

objects from everyday life. This will facilitate the student's encoding of relevant information by relating new entities to familiar, well rehearsed encodings.

4. Simulations should draw the student's attention to aspects of a situation or problem that can be easily dismissed or not observed under normal conditions. The program design should
   (a) promote the simultaneous encoding of key features of the environment,
   (b) make it clear to the student which features of the environment are to be associated with the outcome of a prediction, and
   (c) promote the encoding of information that relates the variability of environmental features to specific actions. (delMas, 1996, p. 8)

These features to promote understanding also represent some of the general best practices listed by EvNet (Table 1). For example, delMas advocates student-centered not instructor-centered instruction, provides student control in actions as well as pacing, and structures the software for meaningful student interaction with the technology. delMas' list of features to promote understanding was developed for statistical simulation-based software but could also be used as a guide for developers of similar software used in other disciplines. The author suggests that the interested reader examine delMas' paper (1996) for complete details.

Current Efforts to Assess Internet-based Instructional Technologies

Digital Libraries offer the promise of classifying and organizing the multitude of instructional materials on the Internet. *Eisenhower National Clearing House*, the *Science, Mathematics, Engineering, and Technology Education* (SMETE) *Digital Library*, and *Multimedia Educational Resource for Learning and Online Teaching* (MERLOT) are a few examples of digital libraries that reach a wide variety of audiences such as K-20 teachers in numerous disciplines. MERLOT, for example, is designed for teachers and students as a source of reviewed online learning materials. The National Science Foundation has supported two digital library projects in the areas of mathematics and statistics. The *Mathematical Sciences Digital Library* (MathDL, www.mathdl.org) is an online resource managed by the Mathematical Association of America and hosted (the library component) by the Math Forum. The site provides online resources for both teachers and students of collegiate mathematics, including the *Journal of Online Mathematics and its Applications* (JOMA, www.joma.org/) a peer-reviewed on-line publication, a catalog of mathematics commercial products that has editorial reviews and reader ratings, and Digital Classroom Resources, which have authors' statements and reader reviews. The Consortium for the Advancement of Undergraduates Statistics Education (CAUSE) has as one of its primary projects the development of a digital library for undergraduate statistics education materials. This digital library, called *CAUSEweb* (www.causeweb.org), includes a peer-reviewed section and a section that organizes and classifies materials existing on the web.

Each of these digital library projects must develop a set of criterion for reviewing and evaluating the potential usefulness of the educational learning materials. CAUSEweb is partnering with MERLOT to learn from their existing review process. Both of these groups use quality, likely effectiveness, and ease of use as their three main evaluation categories. More details of each category follow:

- *Quality of Content*: presents valid concepts, models and skills, presents educationally significant concepts, models and skills for the given discipline
- *Likely Effectiveness as a Teaching-Learning Tool*: presentation of materials improves faculty/student ability to learn material, easily integrated into statistics course/curriculum, teaching/learning goals easily identifiable, conducive for writing good learning assignments, promotes and/or uses effective learning strategies
- *Ease of Use*: easy first time use, consistent appearance, clear instructions, not limited by technical resources, does not need technical support. (Woodard, 2004)

MERLOT's evaluators provide a rating (based on a five star system) to indicate the quality level of the reviewed multimedia materials in their database. In July 2004, CAUSEweb started a peer-reviewed process of assessing undergraduate statistics education learning materials. Users of the CAUSEweb digital library will have access to the reviews of the items selected in the peer-review process.

On a much smaller scale, individuals are developing tools for assessing web-based materials. As one example, at the 2002 Joint Statistical Meetings panel discussion[2] the author presented a rubric for assessing web-based Java applets used in mathematics or statistics classes. In this collaboration with Leann Hooge[3], six main evaluation categories were selected

- Technical Assessment
- Website Assessment
- Text Assessment
- Applet Assessment
- Instructional Assessment
- Mathematical/Statistical Assessment

By developing a series of primarily yes/no questions, we created a form which enables easy evaluation of a website with Java applets. For example, the questions related to Applet Assessment are

- Does the applet allow the student to work at his/her own pace?
- Does the student have control over the amount of information presented?
- Does the student have control over the applet's operation and repetition?
- Does the student have the opportunity to make conjectures and then test them?
- Are the graphics explained for the applet?

---

[2] Our Toolbox is Overflowing: Toward a Framework for Mapping Technologies and Topics in Introductory Statistics."

[3] A graduate student in the Master of Science in Teaching in mathematics at Middle Tennessee State University.

- Does the applet provide information that is interesting and relevant to the student?
- Can the student interpret the results?
- Are the conclusions from the applet obvious to the students?
- Does the applet promote learning or is it just a flashy novelty?

With the exception of the Mathematical/Statistical Assessment category, the majority of the rubric questions are general enough to extend to many disciplines. Answers from all of the assessment questions are compiled into one overall score. Preliminary analysis indicates that experienced teachers can use the tool to evaluate which websites will be useful for student's learning; however, more extensive evaluation is needed. A complete version of this assessment rubric can be found at www.mtsu.edu/~statlist/javarubric.html.

Conclusions

The author believes that there is still much work to be done in developing and disseminating standards for assessing the variety of instructional technologies available for teaching statistics. Assessment instruments and evaluation/review criteria can often be used as a guide for developers of new instructional technologies. The creation of organizations such as the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) provides an important means of bringing together statistics education experts to study one of their self-defined tasks, the complex issues of assessment of educational materials including technology-based materials. Additionally, the development of a digital library of statistics education materials, many of which will be instructional technology materials, offers new opportunities for continued dialog, research, and development of assessment approaches for technology-based materials for teaching statistics. This author hopes that in the next decade additional research will help meet the need Garfield highlighted in 1996 for valid, informative assessment methods for measuring "changes in students' understanding as they interact with technology." (1996, p. 2)

References

Cuneo, Carl. (n.d.). *Twenty Criteria for Evaluating Instructional Technologies: From Best to Worst Practices.* Retrieved July 26, 2004, from http://socserv2.mcmaster.ca/srnet/toolkit/practice/pracindx.htm

delMas, Robert. (1996). *A Framework for the Evaluation of Software for Teaching Statistical Concepts.* Paper presented at the 1996 IASE Roundtable Conference Proceedings. Electronic Version. Retrieved July 15, 2004, from http://www.stat.auckland.ac.nz/~iase/publications/8/7.delMas.pdf

EvNet, Network for the Evaluation of Training and Technology. (n.d.). *About Us: Executive Summary.* Retrieved July 23, 2004, from http://socserv2.socsci.mcmaster.ca/srnet/exsum.htm

Garfield, Joan. (1996). Preface to 1996 IASE Roundtable Conference Proceedings. Electronic Version. Retrieved July 23, 2004, from http://www.stat.auckland.ac.nz/~iase/publications/8/iipreface.pdf

Gilbert, Steven. (n.d.). *What is a Triad?* PowerPoint Presentation. The Teaching, Learning, and Technology Group. Retrieved July 2002, from http://www.tltgroup.org/media/fl/Triad.htm

Gordon, F., and Gorden, S., (Eds.). (1992). *Statistics for the Twenty-First Century*. (MAA Notes Number 26). Washington DC: The Mathematical Association of America.

Moore, T. L., (Ed.). (2000). *Resources for Undergraduate Instructors Teaching Statistics*. (MAA Notes Number 52). Washington DC: The Mathematical Association of America and The American Statistical Association.

Rossman, A. and Short, T. (1996-1999). *STATS: Statistical Thinking with Active Teaching Strategies*. Funded by the UFE program of the National Science Foundation award #DUE 9554621.

Shaughnessy, J. Michael. (1996). *Discussion: Empirical Research On Technology And Teaching Statistics.* Paper presented at the 1996 IASE Roundtable Conference Proceedings. Electronic Version. Retrieved July 23, 2004, from http://www.stat.auckland.ac.nz/~iase/publications/8/17.Shaughnessy.pdf

Siegrist, Kyle. (1997-2004). The Binomial Coin Experiment. *Virtual Laboratories in Probability and Statistics.* Retrieved June 2002, from University of Alabama in Huntsville, Department of Statistics Web site: http://www.math.uah.edu/stat/

Snell, L., Doyle, P., Garfield, J., Moore, T., Peterson, B., and Shah, N. (1999). *Chance.* Funded by NECUSE and NSF DUE. Retrieved June 23, 2001, from Dartmouth College Web site: http://www.dartmouth.edu/~chance/

The Teaching, Learning, and Technology (TLT) Group. (n.d.). Retrieved July 2002, from http://www.tltgroup.org

Valdez, Gilbert. (n.d.). Evaluation Standards and Criteria for Technology Implementation. *North Central Regional Educational Laboratory.* Retrieved July 23, 2004, from http://www.ncrel.org/tandl/eval_standards_and_criteria.htm

Woodard, Roger. (March 19, 2004 and June 2004). Personal Communication of MERLOT and CAUSEweb Peer Review Standards.