The Web-based ARTIST:

Assessment Resource Tools for Improving Statistical Thinking

Joan Garfield and Robert delMas
University of Minnesota
jbg@umn.edu
delma001@umn.edu

Beth Chance
Cal Poly, San Luis Obispo
bchance@calpoly.edu

Paper presented in the Symposium:
Assessment of Statistical Reasoning to Enhance Educational Quality

AERA Annual Meeting
Chicago
April 2003

This project was supported, in part,
by the
National Science Foundation
Opinions expressed are those of the authors
and not necessarily those of the Foundation

Overview

This paper describes the new, NSF-funded Web Assessment Resource Tools for Improving Statistical Thinking (ARTIST) project, which is developing web-based assessment resources for introductory statistics courses. The Web ARTIST project is collecting high quality assessment items and tasks as part of a website that will also provide online testing, offer guidelines for using the assessment tasks, and allow for the collection and compilation of data for research and evaluation purposes. A later goal of the project is a comprehensive test to measure desired outcomes of a first course in statistics.

The Need for Assessment Resources

Inspired by the evaluation of calculus reform courses (Tucker & Leitzel, 1995), a study was conducted by Garfield (2001) to evaluate how the reform movement in statistics education has affected the teaching of introductory statistics courses and how distinctly statistics is taught in different departments and institutions. The results of this study suggest that many statistics instructors are aligning their courses with reform recommendations regarding technology, and to some extent, with teaching methods. Only about one-fifth of the instructors surveyed appear to be teaching a traditional statistics course- relying primarily on the lecture method, not incorporating technology, and using non-reform textbooks. A large percentage of respondents describe changes made in the past few years, with the most frequent changes being in the use of technology, followed by teaching methods and course content. Reform efforts and the increased availability of technology resources appear to be affecting many introductory statistic courses. Most faculty reported positive outcomes regarding changes made: more student satisfaction and increased faculty enjoyment, as well as more sharing of ideas and methods with colleagues.

The weakest area of reform was assessment, where few instructors reported anything other than traditional exams, used only to assign grades. Such assessments mainly require students to recall or recognize definitions, perform calculations, and carry out procedures correctly. This is problematic for two reasons. While some students may earn a good grade in their first statistics course by having good study habits, completing homework on time, and having a good memory and aptitude for mathematics computations and formulas, these students may not be able to actually reason about statistical information or apply what they have learned in other courses or contexts. As more instructors consider students' ability to think, reason, and communicate statistically as desired course goals, different types of assessment methods are needed that focus specifically on these learning goals. In particular, instructors need easy access to a variety of item formats and types of performance assessments that help evaluate student attainment of statistical literacy (e.g., understanding words and symbols, being able to read and interpret graphs and terms), reasoning (e.g., reasoning with statistical information), and thinking (e.g., using statistics to make predictions or judgment in different contexts). Well-designed written assessments can better capture how students think, reason, and apply their learning, but many instructors are possibly intimidated by these methods or are deterred by the amount of time needed to read and evaluate students' responses.

Secondly, in addition to the positive findings of reform recommendations being implemented, and instructors' perceptions of positive outcomes, there is a need for high quality assessments to determine how well the "new" courses actually prepare students to do more than compute, use formulas, and identify correct definitions. Instructors want to determine how well their students are able to demonstrate different aspects of statistical literacy, reasoning, and thinking, and to allow assessment results to inform them where changes are needed in the course curriculum, materials, and use of technology to achieve desired outcomes. For example, many instructors appear to be using new technological resources or new textbooks that reflect reform recommendations for course content. However, assessment materials are not currently available that may be used by faculty to better evaluate the important student outcomes described above. Therefore, it is not known whether using technology and good textbooks is enough, or if there are other important factors that affect student learning (e.g., instructional methods). If high quality assessment materials were available on a website that allowed for customizing quizzes and exams and provided guidelines and examples of other alternative assessments, more faculty could utilize appropriate and informative assessment methods and gather more appropriate information across a variety of courses to support research studies. Without such assessment information, it is unlikely that other instructors will be willing to make these types of changes in their courses.

Traditional assessment methods are primarily used to assign grades and rarely reveal the nature of students' reasoning and thinking. Developing assessments that go beyond traditional methods to evaluate the attainment of statistical literacy, thinking, and reasoning is not a trivial task.

Current Assessment Challenges in Statistics Education

The identification of issues and challenges related to the assessment of student outcomes in statistics courses is a recent development. In their article on assessment in statistics education, Garfield and Gal (1999) identify many of the challenges that need to be addressed. These include:

1. Assessment of "Statistical Literacy"

We need ways to assess the application or transfer of student learning to interpretive or functional tasks such as those encountered in media and outside the classroom.

2. Assessment of Students' Intuitions and Reasoning about Probability and Statistics

We need ways to transfer and adapt promising assessment methods and instruments used by researchers to formats that are reasonably acceptable and accessible to teachers and that can be used for "routine" classroom use.

3. Assessment of Students' Understanding of "Big Ideas" and Statistical Thinking

Assessment items or tasks are needed that can evaluate students' understanding of and sensitivity to the prevalence and importance of the "big ideas" of statistics in different contexts. We also need to assess students' statistical thinking about the nature and process of statistical investigations.

4. Developing Models to Use in Evaluating and Comparing Curricula

As new curricula, innovative textbooks, and instructional software replace traditional approaches to teaching statistics, there is an increasing need for reliable, valid, practical, and accessible assessment instruments to use in evaluating the relative utility of these materials and methods.

The Project

The Web ARTIST project is designed to address these challenges by producing four types of outcomes as described below.

1. A collection of high quality assessment items and tasks, coded according to content and type of cognitive outcome (e.g., literacy, reasoning, or thinking). This will include a variety of item formats including enhanced objective-format questions (e.g., items that require students to match concepts or questions with appropriate explanations) and longer, written assignments such as performance tasks, projects, portfolios, and journals. By the end of the project items will also be given difficulty ratings based on percentages of students who get the item correct.
2. A website that contains the assessment items and tasks, provides guidelines for using the assessment items/tasks in various ways (e.g., online quizzes, offline exams, prototypes of written assignments and scoring guidelines), and allows for the collection and compilation of data for research and evaluation purposes.
3. Faculty development workshops and mini-courses to encourage and assist statistics instructors from various disciplines to use the assessment resources to improve student learning, improve courses, and evaluate course outcomes.

4. A standard, comprehensive test that measures desired outcomes of a first course in statistics that may be used to evaluate a variety of first courses, and allows for comparisons of outcomes across different types of courses (e.g., traditional, lecture-based courses, on-line courses, small interactive classes, etc.)

1. Collection of assessment items/tests/guidelines/examples

We began the project by collecting items to enter in an item bank. These include items developed by the three principal investigators and our advisory board, items from people who volunteered to submit materials that they feel are useful and of a high quality, and items used in relevant research studies. Before beginning the item collection, the project advisors were asked to help generate a list of topics included in a first undergraduate statistics course. After several email discussions we developed the following list of topics to guide the collection and classification of assessment items:

Topics for a First Statistics Course

Data:

Types of data, variables, types of variables such as categorical, quantitative, binary.

Producing and collecting data:

Experiments, observational studies, surveys, samples, population, quality of data, bias, confounding.

Representing data:

Choosing appropriate graphs, constructing and interpreting graphs and tables, summarizing graphs for single variables, both categorical and quantitative, outliers, recognizing shape/trend of data.

Measures of center:

Calculating measures of center, estimating them from graphs, knowing when to use them, how to interpret them, properties.

Measures of spread:

How to find them, how to estimate them from graphs, when to use them, how to interpret them, properties.

Comparing groups:

Comparing two or more groups using graphs and/or numerical summaries.

Measures of position:

Percentiles, quartiles, z scores.

Normal distribution:

Characteristics of the normal distribution, Empirical rule, calculating probabilities, areas under the curve.

Bivariate data, quantitative:

Scatterplots, correlation, simple linear regression, descriptive and inferential methods, outliers, diagnostics, influential observations.

Other types of regression:

Polynomial, nonlinear, multiple regression.

Bivariate data, categorical:

Two-way tables and chi-square test, association.

Probability:

Basic ideas and terms of probability, relative frequencies, random variables, random devices, random number tables/generators, probability rules, simulation.

Binomial distribution:

Binomial variables, probabilities, mean and standard deviation, normal approximation, binomial formula, continuity correction.

Samples and sampling:

Types of samples, sample variability, sampling distributions, Central Limit Theorem.

Hypothesis tests:

Logic of significance tests, one and two sample tests, means and proportions, p-values, types of errors, power, assumptions.

Distributions for test statistics:

Characteristics of the t, F and chi-squared distributions.

Estimation & Confidence intervals:

For one and two sample means and proportions, interpreting confidence intervals, assumptions, margin of error.

One-way analysis of variance

For each topic, the list of categories following each general heading suggests some of the different categories within that topic, but not all. We also realize that some items and tasks may cover more than one topic.

In addition to having items that span the topics listed above, we also need to have items that can be used to assess statistical literacy, reasoning, and thinking. We have reviewed the literature in these areas (e.g., Chance, 2002; Cobb, 1997; delMas, 2002; Gal, 2002; Garfield, 2002; Jones et. al., 2000; Rumsey, 2002; Wild and Pfannkuch, 1999) and have had many rounds of discussions in order to develop descriptions of what we think distinguishes these learning outcomes.

Statistical literacy

Statistical literacy includes basic and important skills that may be used in understanding statistical information or research results. These skills include being able to organize data, construct and display tables, and work with different representations of data. Statistical literacy also includes an understanding of concepts, vocabulary and symbols, and includes an understanding of probability as a measure of uncertainty. (For papers on statistical literacy see Gal, 2001; Rumsey, 2002.)

Statistical reasoning

Statistical reasoning may be defined as the way people reason with statistical ideas and make sense of statistical information. This involves making interpretations based on sets of data, representations of data, or statistical summaries of data. Statistical reasoning may involve connecting one concept to another (e.g., center and spread) or may combine ideas about data and chance. Reasoning means understanding and being able to explain statistical processes and being able to fully interpret statistical results. (For papers on statistical reasoning see Garfield, 2002; Ben Zvi and Garfield, in press.)

Statistical thinking

Statistical thinking involves an understanding of why and how statistical investigations are conducted and the "big ideas" that underlie statistical investigations. These ideas include the omnipresent nature of variation and when and how to use appropriate methods of data analysis such as numerical summaries and visual displays of data. Statistical thinking involves an understanding of the nature of sampling, how we make inferences from samples to populations, and why designed experiments are needed in order to establish causation. It also includes an understanding of how models are used to simulate random phenomena, how data are produced to estimate probabilities, and recognition of how, when, and why existing inferential tools can be used to aid an investigative process. Statistical thinking also includes being able to understand and utilize the context of a problem in forming investigations and drawing conclusions, and recognizing and understanding the entire process (from question posing to data collection to choosing analyses to testing assumptions, etc.). Finally, statistical thinkers are able to critique and evaluate results of a problem solved or a statistical study. (For more papers on statistical thinking see Wild and Pfannkuch, 1999; Chance, 2002.)_

The three types of outcomes described above are all considered goals of reform-based first courses in statistics and stand in contrast to traditional courses that emphasize computations, formulas, and proofs. However, knowing that instructors will also want these types of items, we added a fourth category.

The Role of Computation

Some items that assess statistical literacy, reasoning, and thinking, include some computation, while others do not. A decision was made to only include computational items if they help to assess statistical literacy, reasoning, or thinking and not to include items in the ARTIST data base that only determine if students can compute an answer or use a formula (e.g., calculate the mean of the following set of test scores). Therefore, each item is classified on an additional dimension: includes computation or does not include computation. For example, an item that asks students to determine the standard error for a 95% confidence interval, given the interval endpoints, sample mean, and sample size, involves both computation and reasoning.

Table 1, which expands a table presented by delMas (2002), was used to help classify tasks according to these categories.

Table 1. Tasks that may distinguish item types.

BASIC LITERACY	REASONING	THINKING
WHAT DEFINE IDENTIFY DESCRIBE REPHRASE TRANSLATE INTERPRET READ CONSTRUCT	WHY? HOW? EXPLAIN (THE PROCESS)	APPLY CRITIQUE EVALUATE GENERALIZE

Examples of Items

The following examples may help illustrate the different types of item classifications. We use one problem context to create a variety of item types.

Problem: A random sample of 30 first year students was selected at a University to estimate the average score on a mathematics placement test for all first year students. The average for the sample was found to be 81.7 with a sample standard deviation of 11.45.

Statistical literacy:

- Explain what the standard deviation tells you about the variability of placement scores for this sample.

Statistical reasoning:

- An outlier was found in this data set, with a placement score of 4. If it is removed, how will the mean and standard deviation be affected? Why does removing the outlier have this effect on the mean and standard deviation?

Statistical thinking:

- A psychology professor at a state college has read the results of the university study. The professor wants to know if students at his college are similar to students at the University with respect to their mathematics placement exam scores. This professor collects information for all 53 first year students enrolled this semester in a large section (321 students) of his "Introduction to Psychology" course. Based on this sample, he calculates a 95% confidence interval for the average mathematics placement scores exam to be 69.47 to 75.72. Below are two possible conclusions that the psychology professor might draw.

  1. The average mathematics placement exam score for first year students at the university is lower than the average mathematics placement exam score of first year students at the state college.

  2. The average mathematics placement exam score for the 53 students in this section is lower than the average mathematics placement exam score of first year students at the university.

  For each conclusion, state whether it is valid or invalid. Explain your choice for both statements. Note that it is possible that neither conclusion is valid.

Here is an example using a different problem context:

Problem: The following stemplot displays the average annual snowfall amounts (in inches, with the stems being tens and leaves being ones) for a random sample of 25 American cities:

0	000000024
1	028
2	00228
3	8
4	2248
5	48
6	0

Statistical literacy:

- Describe this distribution.

Statistical reasoning:

- Without doing any calculations, would you expect the mean of the snowfall amounts to be larger, smaller, or about the same as the median? Why?

Statistical thinking:

- A researcher has data on the average snowfall for these same cities from 20 years ago. She wants to test whether snowfall amounts are higher now then they were 20 years ago. Describe an appropriate way to answer this question?

2. The ARTIST Website

Our plans for the actual website include several areas. Our first goal is to develop a useable and flexible structure that allows instructors and researchers from around the country to access items, copy them into a Word file for test construction, or administer them via web-based procedures. We plan to accumulate many student responses in a database archived on the web. This collection of large amounts of classroom-level data from a variety of educational institutions will allow researchers to begin to address some of the most interesting and challenging questions in statistics education.

The website is being designed to meet the following objectives:

1. Make items available to teachers to select/adapt for their courses.
2. Provide online tests for students, with scores reported to teachers and students.

3. Gather data from students to assess the reliability and validity of sets of items (this data collection and analysis will be an on-going activity).
4. Provide tasks and scoring rubrics to assess more complex types of knowledge.
5. Provide additional assessment guidelines related to using assessment to improve student learning, to improve instruction, and to evaluate courses or teaching methods.
6. Provide advice, guidelines, and implementation assistance for instructors (e.g., a section on how to integrate different assessment components into one course).

We currently have a first version of the website that we are using both to share resources and to collect assessment resources (http://www.gen.umn.edu/artist/). PDF files of many articles and chapters on assessment, and links to online articles as well and related websites can be found at the ARTIST website.

The ARTIST Website Design

This section describes our initial conceptualization of the ARTIST website, some of which was influenced by suggestions in the literature (e.g., Wagner, 2001). The functionality of the ARTIST website will be provided through a combination of HTML, JavaScript, VBScript, and Active Server Pages (ASP) programming. The website will consist of several pages that serve as interfaces to the various functions of the site. The homepage for the site will display acknowledgment of NSF funding, a brief description of the ARTIST project, and options for logging on to the website. Once logged onto the site, users are sent to a new page that presents the ARTIST site map.

The following options will be available from the site map:

- Create an assessment instrument.
- Modify an existing assessment instrument.
- Add assessment items to a personal database.

- View results from an assessment.
- Assessment guidelines.

Clicking a button next to one of these options takes the user to a new page.

Creating an Assessment:

The initial page for creating quizzes and exams, called the Item Type page, displays a table of options. Four broad assessment categories of Literacy, Computation, Reasoning, and Thinking, define the columns of the table. The table rows are defined by main topics commonly found in an introductory statistics course, with each topic divided into subtopics. Buttons and lists will be used to guide instructors in selecting items to compose a test. Feedback from the advisory group for this project will be used to define and refine the main topics and subtopics.

Modifying an Existing Assessment Instrument:

Each user will have their own personal set of assessments created from the database of assessment items. When the option to modify an existing assessment is selected from the ARTIST homepage, the user is taken to a page that lists all the assessments they have created. From the Assessment Selection page, a user can choose to modify or delete an entire assessment.

View Results from an Assessment:

This option takes the user to a page that lists all assessments created by the user. When the user selects a particular assessment, a drop-down list consisting of student lists is provided. Once a student list is selected, the user clicks a VIEW REPORT button. This takes the user to an Assessment Report page where she can toggle among three different report formats (Summary, Individual Students, and Item Analysis).

Assessment Guidelines:

This option from the ARTIST homepage takes the user to a page that lists several areas, including:

- Using assessment to evaluate student learning (ways to combine assessment information to provide grades, feedback to students, etc.).
- Using assessment to improve instruction (how to identify areas where students are not achieving the desired learning outcomes).
- Using assessment to evaluate courses or teaching methods (how to analyze student data and compare it to data from other types of courses).
- Using existing scoring rubrics for performance assessments or projects.
- Developing customized scoring rubrics for performance tasks or project.

Materials for each of these areas will be developed from materials already written by the three principal investigators and revised with feedback from the advisory group.

Implementation Issues

We realize that using assessment information involves more than creating a test or choosing an instrument. Several practical issues have arisen as we discussed our project with our advisory group. We now plan to incorporate some questions and answers dealing with these issues in our website as well as include discussions of implementation issues in our faculty development workshops. These issues include:

- Open or closed book exams, and use of notes or sheets of formulas
- Use of technology on assessments
- Use of real data on assessments
- Reviewing before exams
- Going over exams in class or providing correct answers
- In class and take-home exams, and time restrictions

Security Issues

In constructing an item bank for instructors to use to construct tests, and an online testing system, we realize that there are important security issues to deal with. For example, can we hope the database of questions will be large enough so that we do not worry about whether students can access potential test items? One possibility is to make the answers difficult to access, such as locating them in a separate location on the data base. We considered having instructors access the database with a password but realize students could easily obtain passwords as well. We are still working on these issues.

3. Faculty Development Opportunities

An important consideration in developing new tools for statistics instructors is how to encourage instructors to use these tools in appropriate and productive ways. Working with faculty at all stages of the product development is an important part of this project. We have statistics instructors on the advisory group who will advise us on implementation issues and help us make the Web ARTIST appealing and easy to use. Making the assessment tools straightforward to implement and illustrating the utility of the information gained will help more faculty to enhance their own assessment plans.

Mini-courses and workshops will be offered beginning in 2004 to small groups of faculty as they learn how to use our assessment resources in their courses for various purposes. One the most effective ways to encourage faculty use of these tools is through workshops where they can develop, revise, test, and receive feedback on integrating these items for their own formative and summative evaluation purposes. The workshops will also be used to gather formative evaluation for our own use in improving the overall project. A main purpose of these workshops will be to expand awareness of assessment issues among a broader group of faculty.

4. A Common Test for First Year Courses: Comprehensive Assessment of Outcomes in Statistics (CAOS)

Many final exams and standardized tests contain examples of poor statistics and probability questions, which are a misleading reflection of what we want our students to know. For example, multiple-choice items can focus too much on calculation in an artificial setting, with no explanation or interpretation required of the students. A notable exception is the Advanced Placement (AP) Statistics exam for high school students, which serves as a recent example of an assessment tool administered in large groups that attempts to focus on reasoning as well as calculation. However, scoring the exam is a time consuming and complex task.

An important outcome of this project is the development of a set of items to be administered in an online test that can be used in many first year courses to evaluate the attainment of desired student outcomes. Working with our advisory group and other colleagues teaching statistics, we will develop a common set of items to be used across different courses and institutions. These items will represent the big ideas, the dispositions, and the types of reasoning, thinking and literacy skills deemed important for students across first courses in statistics. Early versions of the instrument will be piloted in the second year of the project and data will be gathered on a wide scale during the third year of the project. Information will be collected on the type of textbook, teaching methods, and technology used at each institution to examine relationships between outcomes and instructional variables.

Measurement Issues

We have two intended goals for our materials and these goals lead directly to selection of appropriate methods for calibrating and determining the quality of items. One purpose is for use by teachers for assessing their students' learning in a course. This would be a criterion-referenced setting where teachers are interested in determining students' mastery of material and achievement of desired learning outcomes. In this case, traditional estimates of reliability are not appropriate because it applies to a norm referenced group and is also sample specific (i.e., dependent on the ability range in the sample).

Determining measurement precision and accuracy of classifying students (mastery vs. non-mastery) is relevant to criterion-referenced testing. In general, precision is more affected by the number of items used in an assessment. Therefore, we will need to identify the number of items needed to achieve a certain level of accuracy for assessing a particular topic or learning outcome. We will also try to gather inter-rater reliability for scoring rubrics for performance assessment tasks.

For forced choice questions, Rasch models can estimate measurement error for estimating various ability levels and can be determined with samples of size 30 or more. The Rasch model estimates item difficulty (e.g., ability level at which a student has a 50% chance of getting an item correct) and, as such, are mathematically sample independent. Rasch ability estimates are logits in the range of –3 to +3 that can be rescaled. We plan to use the Rasch model to determine item difficulties (for more on Rasch measurement, see Wright and Stone, 1979). This information will be used to set up a matrix that presents topics on one dimension, type of learning outcome (Literacy, Calculation, Reasoning and Thinking) as a second dimension, and then, for items in each cell, have the item difficulty as a third dimension.

Eventually, we plan to suggest ways for instructors to build tests utilizing these three dimensions to create tests that reduce measurement error to an acceptable level. For example, to build criterion-referenced tests that are keyed to a course, the selected items need to match the course content and desired cognitive outcomes, and represent the range of abilities of students (item difficulty).

The other purpose of the assessment materials is for use in research and evaluation studies. For these purposes, subsets of items will be constructed for particular topics or outcomes and offered as scales, in addition to the CAOS test. These scales will be tested for reliability and modified to have high internal consistency reliability coefficients. Reliability is sample specific, so identification of meaningful samples of students in which to gather data will be a key issue.

For analysis of validity, we plan to use expert judges. Our advisory board will be asked to rate items in terms of content and outcome in order to calculate percent agreement between raters. We believe that this type of content validity is appropriate for classroom testing use. For research use, construct validity analyses may be used to demonstrate that test performance is related to independent measures of constructs. For example, we might try to examine correlations with AP Statistics test performance or grades in a statistics course. The sensitivity of ARTIST test to instruction will be investigated by determining if test performance discriminates between students who have and have not had instruction on a topic. Sensitivity to type of instruction might also be demonstrated. Finally, we plan to use some think-aloud protocols with small groups of students to determine construct validity.

Class Testing of Assessment Items

There will be two levels of class testing. The first set of class testing will be more informal. This will involve people who have signed up via our website to be class testers. They will be instructed to select whatever items they want to use. They will use these items and offer feedback on wording, student perceptions, and their own reactions to the items or materials.

A second level of class testing will be more structured, where selected instructors are invited to test particular scales or the entire CAOS test. We will try to control the testing conditions and will try to gather some additional information to use in our analyses of validity.

Project Impact

We believe that the ARTIST project has the potential to improve the teaching and learning of statistics in first courses by providing a freely available and much-needed resource for course instructors. This resource will allow for the collection and analysis of student assessment data in statistics courses across departments and institutions. It will provide a centralized database to use in future evaluation and research studies, resulting in detailed information on best practices of teaching statistics and identifying factors related to the attainment of statistical literacy, reasoning, and thinking.

We are fortunate to have an excellent and helpful set of advisors that have been helping us work toward achieving these goals. The advisory group consists of a variety of faculty who represent different educational settings (departments and institutions) as well as areas of expertise (e.g., measurement, evaluation, attitudes, statistics instruction), and have experience working with diverse populations of students. The advisory group includes:

Julia Clark – Mathematics, Hollins University
George W. Cobb – Mathematics, Mount Holyoke College
John P. Holcomb, Jr. – Mathematics, Cleveland State University
Frances Lawrenz – Educational Psychology, University of Minnesota
Carl Lee – Mathematics, Central Michigan University
Anthony Onwuegbuzie –Educational Leadership, Howard University
Roxy Peck – Statistics, California Polytechnic State University
Michael Rodriguez – Educational Psychology, University of Minnesota
Allan Rossman – Statistics, California Polytechnic State University
Deborah J. Rumsey – Statistics, Ohio State University
Candace Schau – Educational Psychology, CS Consultants

Future Plans

Although we are only half way through our first year of the project, it is already clear that many more things could be done beyond the scope of the proposed project. A short list includes having self-assessment capabilities on our website, developing a diagnostic pretest for first courses in statistics, creating a bulletin board on our website for discussions of assessment issues, and developing an electronic newsletter for users of the site. We also have ideas for future work on using assessment to improve instruction by establishing collaborative groups of faculty to build lessons designed to achieve assessment outcomes. Therefore, we are considering a future proposal to build on and extend our current work. We invite interested participants to contact us.

References

Chance, B. L. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment. *Journal of Statistics Education* [Online], *10*(3).

delMas, R. C. (2002). Statistical Literacy, Reasoning, and Learning: A Commentary. *Journal of Statistics Education* [Online], *10*(3).

Cobb, G. (1993). Reconsidering Statistics Education: A National Science Foundation Conference, *Journal of Statistics Education* [Online], *1*(1).

Cobb, G. W. (1997). Mere Literacy is Not Enough. In L. A. Steen (Ed.), *Why numbers count: Quantitative literacy for tomorrow's America* (pp. 75-90). New York: College Entrance Examination Board.

Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review*, *70*, 1-51.

Garfield, J. (2002). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education* [Online], *10*(3).

Garfield, J. (2001). Evaluating the Impact of Educational Reform in Statistics: A Survey of Introductory Statistics Courses. Final Report for NSF Grant REC-9732404.

Garfield, J. and Gal, I. (1999). Assessment and Statistics Education: Current Challenges and Directions. *International Statistical Review*, *67*, 1-12.

Garfield, J., Hogg, B., Schau, C., and Whittinghill, D. (2002). First Courses in Statistical Science: The Status of Educational Reform Efforts, *Journal of Statistics Education* [Online], *10*(2).

Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. J. (2000). A Framework for Characterizing Children's Statistical Thinking, *Mathematical Thinking and Learning, 2*, 269-307.

Hogg, R. (1992). Report of Workshop on Statistics Education, in *Heeding the Call for Change: Suggestions for Curricular Action*, MAA Notes Vol 22, 34-43.

Moore, D.S. (1997a). New Pedagogy and New Content: The Case of Statistics. *International Statistical Review, 65*, 123-137.

Moore, D.S. (1997b). Response. *International Statistical Review, 65*, 162-165.

Performance Assessment Links in Science (PALS) (1999). SRI International, Center for Technology in Learning. [Online]. http://pals.sri.com/

Rumsey, D. J. (2002). Statistical Literacy as a Goal for Introductory Statistics Courses. *Journal of Statistics Education* [Online], *10*(3).

Scheaffer, R.L. (1997). Discussion. *International Statistical Review, 65*, 156-158.

Tucker, A. and Leitzel, J. (Eds.) (1995). *Assessing Calculus Reform Efforts.* Washington, D.C.: Mathematical Association of America.

Wagner, E. P. (2001). Customized Internet Assessments. *Journal of College Science Teaching*, *30*(7), 454-476.

Watkins, A. and Loftsgaarden, D. (1999). The CBMS Survey and Statistics Teaching in the United States. Paper presented at the Joint Statistics Meetings, Baltimore.

Wild, C. J., and Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review*, *67*, 223-265.

Wright, B. D., and Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago, IL: MESA Press, University of Chicago.