

ASSESSING STUDENTS' STATISTICAL REASONING

Robert delMas, Ann Ooms, Joan Garfield
University of Minnesota, United States

Beth Chance
California Polytechnic State University, United States
delma001@umn.edu

This paper describes the ARTIST project which was designed to address the assessment challenge in statistics education. The goals of the ARTIST project are to assist faculty who teach statistics across many disciplines in assessing student learning of statistics, enabling them to better evaluate individual student achievement, to evaluate and improve their courses, and to allow them to assess the impact of reform-based instructional methods on the attainment of statistical literacy, reasoning, and thinking. ARTIST consists of a website that provides resources designed to meet these goals. Among the resources are a large, searchable assessment item database, several online topic tests, and a comprehensive test of statistical literacy and reasoning (CAOS). Details of the development of the ARTIST resources, results from an extensive evaluation of the project, and the development of future ARTIST resources are presented.

THE ARTIST PROJECT

The National Science Foundation (NSF) funded the Assessment Resource Tools for Improving Statistical Thinking (ARTIST) project (DUE-0206571) to address the assessment challenge in statistics education as presented by Garfield and Gal (1999), who outlined the need to develop reliable, valid, practical, and accessible assessment instruments. The ARTIST website (<https://app.gen.umn.edu/artist/>) now provides resources for evaluating students' 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., asking questions and making decisions involving statistical information). These resources were designed to assist faculty who teach statistics, across various disciplines (e.g., mathematics, statistics, and psychology), in assessing student learning of statistics, to better evaluate individual student achievement, to evaluate and improve their courses, and to assess the impact of reform-based instructional methods on important learning outcomes. The project is run by a team of three co-investigators (delMas, Garfield, and Chance), each with unique areas of expertise in statistics education, and a graduate research assistant (Ooms, now a post doctoral fellow) with expertise in technology and evaluation. The ARTIST project has been fortunate to have a strong and diverse advisory group: Julie Clark (Hollins University), George W. Cobb (Mount Holyoke College), John Holcomb (Cleveland State University), Frances Lawrenz, (University of Minnesota), Carl Lee (Central Michigan University), Marsha Lovett (Carnegie Mellon University), Anthony Onwuegbuzie (University of South Florida), Roxy Peck (California Polytechnic State University), Michael Rodriguez (University of Minnesota), Allan Rossman (California Polytechnic State University), Deborah J. Rumsey (Ohio State University), and Candace Schau (CS Consultants). The advisors have provided expertise in development, evaluation, and implementation of assessment items, resources, and instruments.

By the beginning of its fourth year (Fall, 2005) the project had produced the following products:

- A collection of over one thousand high quality assessment items and tasks, coded according to content (e.g., normal distribution, measures of center, bivariate data), type of cognitive outcome (e.g., statistical literacy, reasoning, or thinking), and type of item. Users can use a set of linked pages (called the Assessment Builder) to search, review, select, and download items into rtf formatted files that may be saved and modified on their own computers with a word processing program.
- A website that provides access to the assessment item database, as well as many other resources (e.g., references and links to articles on assessment, information on alternative assessment methods including samples of project guidelines and student work, grading

rubrics, research instruments, materials from professional development offerings, ARTIST advisory board responses to questions on assessment implementation issues, web links, etc).

- Eleven online unit tests that measure conceptual understanding in important areas of a first course in statistics which have high validity and reliability.
- A Comprehensive Assessment of Outcomes in Statistics (CAOS) that measures basic statistical literacy and reasoning.
- Several mini-courses, workshops, and conference presentations to encourage and assist statistics instructors in how to use assessment resources to improve student learning, improve their courses, and evaluate course outcomes.

DEVELOPMENT OF ARTIST ASSESSMENTS

The assessment item database was one of the first products to be developed. Items were initially pulled from exams of the project staff (co-investigators and advisory board members) and also solicited from the statistics community through a posting on the ARTIST website. These items were then reviewed and organized by topic and learning outcome. The co-investigators and some of the advisors reviewed items to organize them by topic and learning outcome. Items that were purely computational were eliminated from the database, unless they could be modified into a literacy, reasoning, or thinking type of item. A context was developed and added to items that were not set in a context. Most true/false items were changed into three-option forced-choice items, and forced-choice versions were created for numerous open-ended items (although there are also many open-ended items in the database). All items were edited for statistical content and typographical errors. Knowing that some errors may have been missed, a mechanism was developed so that users of the assessment item database can report concerns with individual items. The database currently consists of more than 1100 items, with new items being added periodically as they are submitted by ARTIST users and reviewed.

The eleven online unit tests and CAOS were developed through a process that took over two years. During this process the ARTIST team developed and revised items and the ARTIST advisory board provided valuable feedback as well as rating the validity of items, which was used to determine and improve scale validity. The topics for the 11 online unit tests are: Data Collection, Data Representation, Measures of Center, Measures of Spread, Normal Distribution, Probability, Bivariate Quantitative Data, Bivariate Categorical Data, Sampling Distributions, Confidence Intervals, and Tests of Significance. These tests cover an intersection of topics included in most introductory statistics courses. Each test consists of seven to twelve multiple-choice items that assess literacy and reasoning for that topic. Online versions of each topic scale were created and evaluated in two rounds of class testing and test revision as described below.

The CAOS test was developed through a similar process of development, revisions, feedback from advisors and class testers and a large validity assessment using 30 experienced Advanced Placement Statistics readers. The current version of CAOS consists of 40 multiple choice items and can be administered online or in a print copy using a machine scannable bubble sheet. Topics covered on the CAOS test assess basic literacy and reasoning about descriptive statistics, probability, bivariate data, and basic types of statistical inference. Again, the intent was to develop a set of items that students completing any introductory statistics course would be expected to understand.

In order to access the online tests, an instructor requests an access code, which is then used by students when they are ready to take the test. As soon as the students have completed the test, either in class or out of class, the instructor may download two reports of students' data. One is a copy of the test, with percentages filled in for each response given by students, and with the correct answers highlighted. The other report is a spreadsheet with percentage correct scores for each student.

CLASS TESTING

A large scale class testing of the online instruments was conducted during spring 2005. Students for the study were obtained through invitations sent to high school Advanced Placement (AP) and college statistics instructors through email lists of major United States organizations

likely to have a membership that represents this population (e.g., AP listserv, Statistics Education Section of the American Statistics Association), ads placed in magazines and newsletters (e.g., AMSTAT news), and information posted on the ARTIST website. Instructors registered their students to take the ARTIST topic scales at points in their courses when students had covered the material assessed by a scale. Nearly 100 secondary-level students and 800 college-level students participated. The spring 2005 results were used to make minor revisions and produce final versions of each scale during summer 2005.

During the class testing phases, data were gathered and used to review responses to each item so that responses could be revised or deleted as needed. Input from class testers was also utilized during this period and used to revise and improve the tests. Class testers have found these tests to be useful for a variety of purposes: testing (for a grade), review, self-testing, and extra credit.

The eleven topic scales and the CAOS test were administered in a second large scale testing during fall 2005 and spring 2006. Final results of this large scale, national sample of secondary and college-level students will be reported in a later paper.

EVALUATION OF THE ARTIST WEBSITE

After developing the ARTIST website, item data base, and online assessments, we wanted to know how instructors were using these materials and how the materials were helping to achieve the original project goals related to improving the teaching and learning of statistics. The NSF granted supplemental funds to conduct a large scale evaluation of the ARTIST project.

The second author created an evaluation model called The Iterative Evaluation Model for Improving Online Educational Resources as part of her doctoral thesis (Ooms, 2005). The model consists of four components (web design, web content, use of the materials, and impact on teachers and students). The model was used as a guide to select data collection methods and to create the data collection instruments. The evaluation of the ARTIST website included:

- A 50-item online survey of ARTIST users to learn about how they were using the materials and how they were changing their assessment practices. Ninety-eight ARTIST users responded to the online survey.
- A 52-item online survey of non-users to find out why they were not using our resources and if they differed from the users in important ways, such as wanting more computational items. The survey was identical to the one used for ARTIST users except for the addition of two questions. Eighty-nine non-users responded.
- Observations of two new and three experienced users of the ARTIST assessment item database as they interacted with web pages designed to access the assessment item database. Participants were asked to think-aloud while performing certain tasks, such as creating a test, adding a question to a test, and removing a question.
- Interviews with seven non-users of ARTIST resources to find out why they were not using these materials.
- Interviews with seven frequent users of ARTIST resources to learn about how their use of ARTIST materials and resources had impacted their assessment practices and/or teaching.

Results were summarized and examined for the four components of the evaluation model. In terms of web content, all the evaluation findings were positive and did not indicate the need for major changes to the website. The user survey respondents were fairly regular users of the ARTIST website, with 85% having visited the site more than twice, and 54% having visited the website more than 5 times. The website seems to be reliable, with less than 14% of the people who browsed the website indicating they encountered at least one error message. With respect to navigating the ARTIST website, 94% reported that the links were labeled in a way that was descriptive of their content. The organizational scheme of the website received high ratings: 74% rated the quality of the organizational scheme as good, and 19% rated it to be excellent. Ninety percent of respondents found it easy to locate a particular topic within the website. The majority of respondents (more than 90%) reported finding the website “somewhat attractive” to “attractive.”

Results from the non-user survey indicated that many had not used ARTIST resources because they were not aware of the website. Forty-five percent of the non-users had not heard of the ARTIST website, and 69% were not aware of the ARTIST assessment item database resource. Interviews with non-users identified some misconceptions about the ARTIST website (e.g., it was for the assessment of teaching, or that it required students to have online access in the classroom). As a result, we have publicized the website and its resources more broadly with detailed information about what it provides.

In terms of web design, survey results showed that some steps in the process of using the web resource were “somewhat difficult to use.” A summary of the results from questions on ease of use for the Assessment Builder indicated that users found it fairly easy to log on, to view items, select and remove items, and to download a set of selected items. More difficulty was indicated for conducting and refining a search. Respondents also rated the quality of several features of the Assessment Builder (e.g., instructions, information buttons, navigation, response time). All features were rated to have high to very high quality by a majority of the respondents. Observations of new and experienced ARTIST users identified problem areas which suggested many changes and improvements in the web design and content that have been made or are currently being made (e.g., ways to improve searching the assessment item database; the addition of large data sets for use in take home finals or student projects).

Evaluation data was also gathered on estimated frequency of use of various ARTIST web pages and ratings of their usefulness. Some web pages were clearly being used more than others. Of the web pages that were rarely used, the vast majority of those who accessed these pages still found them useful. One question that the evaluation was not able to address was why people who visited the website once did not return to use it again.

The section on the Assessment Builder in the ARTIST user survey contained a set of questions on the quality and use of the assessment items. Respondents were asked to compare the quality of the ARTIST assessment items to items from other sources (e.g., textbook item databanks). In general, ARTIST items were found to be of the same or higher quality. Notably, ARTIST items were judged to assess more for conceptual understanding and to incorporate a context more often when compared to items from other sources. Ninety-one percent of the respondents used the ARTIST items in quizzes and for exams. The second most frequent use of the items was for review (44%), and the third most frequent use was for class examples (37%). Twenty-five percent of respondents used the items for class group activities and 19% for homework.

Evaluation of educational impact of ARTIST products was based on interviews with ARTIST users. At the time of the interviews, ARTIST users had not implemented any changes to their instruction. However, there was evidence that they had started to think about changes to their teaching, and that the ARTIST products were having a positive impact on instructors’ perceptions. The survey and interview results indicate that instructors who use many of the ARTIST scales and tests have been rethinking some of their teaching approaches in order to have a greater impact on students’ statistical reasoning and thinking. Still, it was not possible at this stage to evaluate the educational impact of ARTIST in terms of the educational environment.

FUTURE DEVELOPMENTS

Work is currently being done to provide large data sets for use by instructors for assessment purposes. These data sets consist mostly of data collected on students (e.g., body measurements, first day survey data, admissions data) and can be used by instructors to have students solve more open ended, exploratory problems using data and statistical software for assessment purposes. Instructors will be able to download entire data sets, or to select random subsets of records which could be used to give each student in a course a different data set for conducting an out-of-class project.

There are plans to add three instruments to the ARTIST website in coming years.

1. Student Attitudes Towards Statistics (SATS): An online version of the Student Attitudes Towards Statistics (Shau, 1994), a 36 item test that assesses six components of attitudes toward statistics (affect, cognitive competence, value, difficulty, interest, and effort).

2. Statistical Thinking Assessment (STA): A 15-item multiple choice instrument to assess students' statistical thinking at the end of an introductory course. This instrument assesses students' ability to select appropriate procedures, to check conditions and assumptions, and to draw correct conclusions.
3. Statistics Teaching Inventory (STI): This instrument will be developed to gather information on teachers' own beliefs about teaching, the alignment of their instructional practices with current reform recommendations, and the constraints in which they teach (school and student variables). This instrument will be given to instructors whose students take tests on ARTIST, and the data will be stored so that it may be analyzed along with the student test data.

As part of a new effort by the Consortium for Advancing Undergraduate Statistics Education (CAUSE), plans are being made for a consortium of institutions to jointly collect data using the ARTIST instruments to address several questions of interest, such as:

- What is the relationship between students' statistical thinking, literacy and reasoning?
- How are different teaching techniques related to students' attitudes and learning outcomes?
- What is the relationship between students' attitudes toward statistics and their achievement?
- How do attitudes change (or not change) across statistics courses?
- Do changes in student attitudes correspond to learning gains?

The addition of the SATS, STA, and STI to the ARTIST website will allow us to extend our understanding of the roles that student attitudes and motivation play in the learning of statistics. We plan to work closely with current and future projects run by CAUSE to connect their professional development workshops and activities with our data collection efforts. In this way, we may explore the impact of the professional development activities on student attitudes and learning outcomes as well as on teaching variables.

REFERENCES

- Cobb, G. (1992). Teaching statistics. In L. Steen (ed.), *Heeding the Call for Change: Suggestions for Curricular Action*, (pp. 3-33), MAA Notes #22. Washington, DC: Mathematical Association of America,
- Consortium for the Advancement of Undergraduate Statistics Education (CAUSE), <http://www.causeweb.org/>.
- 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.
- Ooms, A. (2005). *The Iterative Evaluation Model for Improving Online Educational Resources*. Doctoral Thesis, University of Minnesota.
- Schau, C. (1994). *Survey of Attitudes Toward Statistics* (Revised) [SATS]. ETS Tests in microfiche collection: 019712.