

# **A Model to Evaluate Online Educational Resources in Statistics**

## **1. INTRODUCTION**

As Internet access in educational settings expands (Kleiner & Lewis, 2003), educators' interest in using online educational resources has steadily increased. In statistics education, internet resources include collections of activities, data sets, web applet, text materials and assessment items. Websites may present rich and motivating materials, but may also contain misinformation or biased information, incomplete instructions, contradictions, or out-of-date material. The inconsistent quality of internet resources may be partly due to the lack of a review mechanism to determine the credibility, quality, and accuracy of published websites (Branch, Kim, & Koenecke, 1999).

In their 2000 Report of the Web-Based Education Commission to the President and the Congress of the United States, there was a call to: "develop high quality online educational content that meets the highest standards of educational excellence" (p. iv). The report stated: "Challenges await content developers and educators in producing, distributing, cataloging, indexing, and evaluating good online content" (p. 69).

In recent years, several evaluations of online educational resources have been conducted (e.g., Fitzgerald, Lovin & Branch, 2003; Jenkins & Robin, 2002; Marchionini, 2000; Branch, Kim & Koenecke, 1999). However, missing from these evaluations was use of a validated, theoretically based evaluation model that would provide guidelines and methods to evaluate the educational impact of online educational resources. This paper describes such a model and an example of its practical application. Since the term model has a different meaning in an evaluation context than in a statistical context, we first provide some explanation about the meaning and utility of an evaluation model.

### **1.1 Evaluation Models**

To a statistician, models provide a way to describe how something works in the real world. Statisticians use statistical models to describe, predict and explain phenomena. In contrast, in the field of evaluation, an "Evaluation Model" is an essential vocabulary term used in educational and program evaluation planning. In this context "model" represents an approach used to guide the design and implementation of a particular evaluation and leads to the use of methods and development of questions.

There are a wide variety of evaluation models used to guide evaluators in planning and conducting a particular type of evaluation (see Stufflebeam and Shinkfield, 2007, Fitzpatrick, Sanders, & Worthen, 2004). House (1978), created a taxonomy of evaluation models that were geared to different types of context and desired outcomes. For example, the widely-used Context, Input, Process, Product evaluation (CIPP) model (Stufflebeam,

2000) is used to guide management-oriented evaluations. Evaluations designed using this approach focus on just one or a combination of the four strands, context, input, process or output of a program being evaluated. There is definitely some overlap between the different evaluation models and it is the task of the evaluator to select the most appropriate approach, while considering the program under evaluation and the evaluation aims.

Among the broad array of evaluation models some are better suited for particular types of evaluations, and professional evaluators tend to prefer one type of model to use in their work. While most of these models may be used to conceptualize and design an evaluation of online educational materials, it would be up to the evaluator to consider all the different aspects of the resource that would need to be included, and not all aspects might be included in different evaluations. Therefore, having a model that is unique to evaluating online educational resources seemed to be something that would aid educational evaluators and fill a need in the evaluation field (Ooms, 2005).

In the field of statistics education, an increasing number of online resources have been designed to improve the teaching and learning of statistics (see Chance et al., 2007). Many of these resources can be accessed through the website CAUSE ([causeweb.org](http://causeweb.org)). Agencies like the National Science Foundation (NSF) have funded many new online resources designed to make the teaching and learning of statistics more effective, particularly at the postsecondary level. An evaluation model that could be used to evaluate and improve existing educational resources is needed. Therefore, the NSF-funded ARTIST project received funding to create, validate, and use such a model. This led to the design and demonstrated use and value of the IEM (Ooms, 2005).

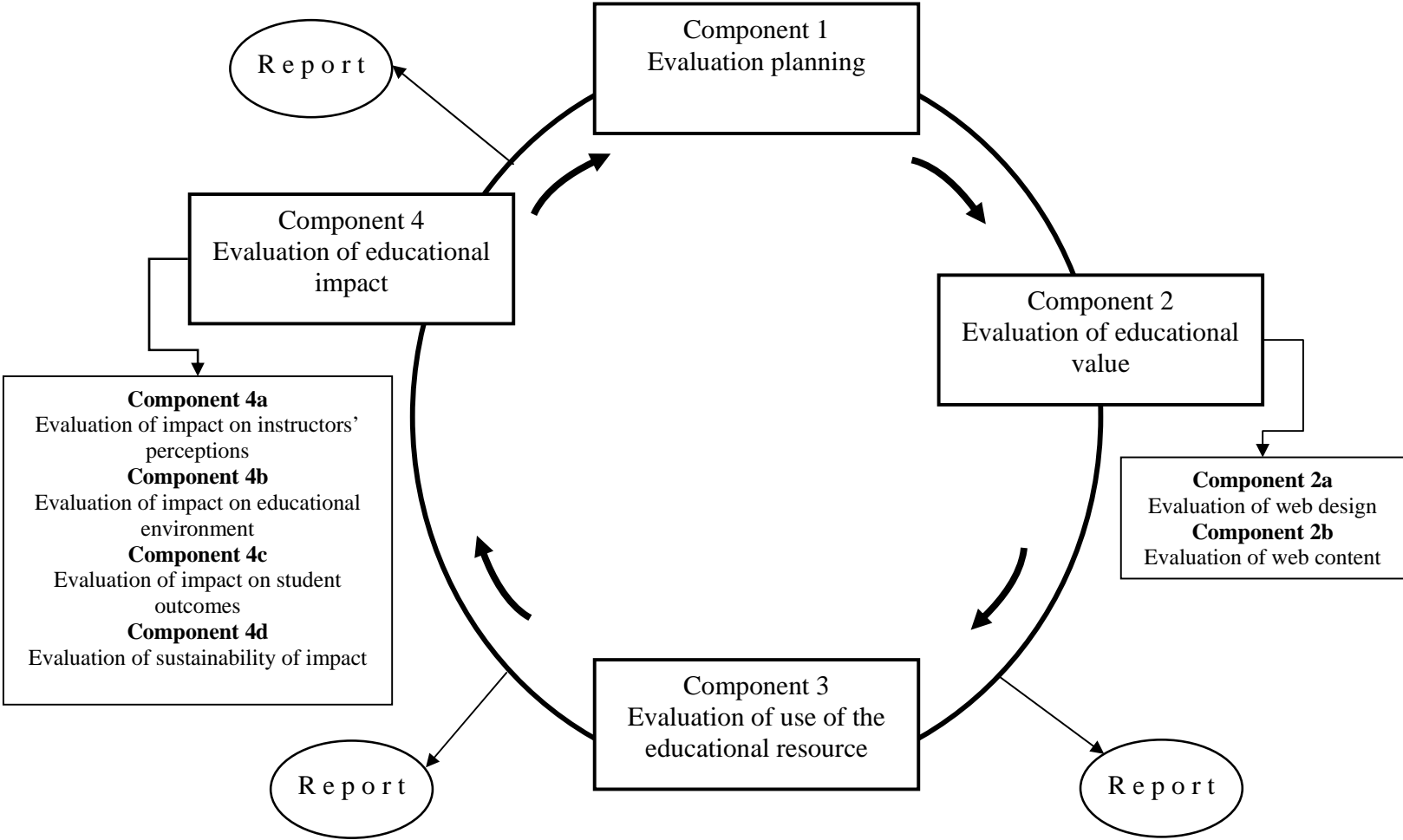
## 2. THE ITERATIVE EVALUATION MODEL FOR IMPROVING ONLINE EDUCATIONAL RESOURCES

The IEM was designed to provide a model to guide professional evaluators with the evaluation of online educational resources and may be used by developers of online educational resources working, with an evaluator, to collect and provide feedback on the use and value of these resources. As an evaluation model, its purpose is to provide evaluators with information to improve a resource. However, the model is not intended to be used by individual educators when they are selecting educational resources. The ultimate aim of the evaluation is the improvement of the online educational resource (formative evaluation), as well as a judgment of the value of the resource (summative evaluation).

Four components are described by the IEM. They are: evaluation planning, web design and content, use of the educational resource, and educational impact. To validate the model, the evidence-based approach to validity was used, where one collects evidence to build the validity argument (AERA, APA, NCME, 1999; Kane, 2001; Messick, 1989; Moss, 1995; Sax, 1997). Evidence from seven different sources was used to build the argument for the validity-based development of the model: (a) a thorough review of the literature, (b) feedback from internal experts, (c) feedback from external experts, (d) field testing of the model with feedback of the users of the evaluation report, (e) evaluation of the evaluation model by the users of the evaluation report, (f) a reflective process by evaluator, and (g) validation by external experts. As you can see in point (d), the field testing of the model is part of the validation process of the model. Therefore, the evaluation model and the evaluation of the ARTIST online resource were developed in an iterative way. The model was developed first but later revised, based upon findings from its pilot-testing. Pilot-testing meant the evaluation model put into practice while evaluation the ARTIST resources. The methodology, the evidence-based approach to develop and validate evaluation models, is described in detail in Ooms, 2005. The IEM has four components, as shown in Figure 1. These components correspond to the four phases of evaluation of an online educational resource. They are 1) Evaluation planning, 2) Evaluation of educational value (evaluation of web design and web content), 3) Evaluation of use of the educational resource, and 4) Evaluation of educational impact (evaluation of impact on instructors' perceptions, on educational environment, on student outcomes, and of sustainability of impact). The figure shows the iterative nature of the model, how each cycle feeds into a repetition of the cycle, and also how results are reported for each component and feed into the next stage of the evaluation.

The following sections provide details of the four phases or components of the IEM. For each component, tasks, evaluation questions, justifications, recommended best methods, and sample questions, are provided.

**Figure 1: The Iterative Evaluation Model for Improving Online Educational Resources**



## 2.1 Evaluation Planning (Component 1)

As in every evaluation, the IEM includes a thorough evaluation planning phase. There are two guiding principles for this first component of the IEM. One is that *The Standards for Educational Evaluation (1994)* be used to plan the evaluation. These standards provide detailed guidelines on evaluation utility, feasibility, propriety, and accuracy. In other words, they offer standards that evaluators should adhere to in order to make their evaluations useful, feasible, and accurate. The second guiding principle is that the evaluation planning needs to be flexible so that it can address evolving questions and address changes that occur in the content or use of the resource being evaluated. This extends to the development of evaluation instruments so that information collected from one instrument can be used to the development of subsequent instruments and items. This first planning phase is not unique to the IEM but builds on principles and activities in other models (Fitzpatrick, J., Sanders, J., & Worthen, B., 2004). The main planning activities are described below.

First, good working relationships should be established with the people who create and maintain the online resources. In addition, it is important to identify and include stakeholders in the evaluation plans. Stakeholders (organizations or people who have a direct interest in the resource) provide important information about the purpose of the website, the intended audience, and they can also assist with establishing criteria/standards of the online resource.

Intended users as well as non-users of the online resource need to be identified, some of whom may be invited to provide feedback during the evaluation process. They will provide essential information. Content experts also need to be identified. They may be used during the evaluation process to provide information about the content of the website. External content experts that may have different backgrounds and perspectives should be invited to participate in the evaluation.

Finally, a needs assessment is conducted during this planning phase to help the evaluator learn about the needs of the intended audience and collect additional information about the intended users. This information may include (1) availability of technology tools for the users (hardware and software), (2) users' level of technology skills, (3) their type of position (e.g., teacher, administrator), (4) level of institution (ES, MS, HS, Higher Ed), and (5) their content area. The alignment of the characteristics of the resource with the characteristics of the users also needs to be evaluated.

## 2.2 Evaluation of educational value (Component 2)

There are two important areas to examine in determining the value of online resources. They are web design and web content.

### **2.2.1 Evaluation of Web Design**

In evaluating the web design, the IEM specifies that the evaluator observes both first time users and regular users while using the web site. This may involve examining users who are using different hardware, software, and internet connections. Users can be asked to perform certain tasks, as they are observed, to see how much time it takes them, and how they feel about performing that task (by asking them to talk out loud as they use the site). Additional activities in this component include a survey of users of the website, and opinions of experts (those who have expertise both in web design and the content area).

A series of questions regarding web design are focused on the accessibility of the website for intended users. This is important because the online resource will not be used if users have no access or slow access to it. Site access depends on the availability of technology tools (hardware and software) for the intended audience. Questions to use include:

1. Is the resource easy to find by using an internet search engine?
2. Did user encounter error messages?
3. How long does it take to open each webpage?
4. How long does it take before all the graphics have been downloaded on each page?
5. Ask users about their level of satisfaction with the accessibility of the webpage.
6. Ask users if they have access to the website and if that access is sufficiently fast according to their expectations.
7. Is the website accessible for auditory or visually impaired users?

A second area of focus is the navigation of the website. A highly navigatable website will be visited more often. Users will be more motivated to return and to spend more time on the website. The navigatable quality often depends on the level of technology skills required to navigate within the website. Questions that address the website navigation ease include:

1. Is the information structure logical?
2. Is it easy to navigate within the website?
3. Are the links clearly and logically labeled?
4. Are the links labeled in a way that is descriptive of their content? If not, which ones are not and why?
5. How logical is the organizational scheme of the website? If not logically structured, what needs improvement?
6. Is it easy to locate a particular topic within the website? If not, what and why?
7. Is it easy to find what they are looking for? If not, what and why?
8. Ask users to rate the difficulty of each of the components involved in using the website and list the components. Which components are difficult, and why?

There are also additional aspects of the web design to evaluate, such as the aesthetic appeal of the website. Aspects such as balance of content, fonts, colors, and sound effects can be inviting or discouraging. To evaluate these characteristics, web users may be asked if they find the website appealing or not, and why. Another aspect is whether or not there is a help function and whether it works efficiently. Here, users who have used the help function can be asked to rate their level of satisfaction with the help function; if not satisfied, why not? If there is a site index, users can be asked to rate their level of satisfaction with the site index in terms of accuracy, in terms of completeness, and in terms of organizational scheme; if not satisfied, why not?

### **2.2.2 Evaluation of Web Content**

To evaluate the web content of the educational resource, data is collected from content experts as well as users. This can be done by surveys and observations. Some areas of web content to examine are whether the goals and intended use are clearly stated, if the content is relevant to the goals, and if the purpose of the website is reflected in the content. The scope of the content also needs to be examined, which should cover the appropriate domains in adequate depth and breadth. The content needs to be complete and resources need to be provided where the content is incomplete. Users and content experts may have different opinions about this. Questions to use in evaluating the scope of web content include:

1. Is the content biased? Why? If not, what is biased?
2. Is the information available covering the domain in an appropriate way? Why? If not, what is missing?
3. Is the content of the website providing completeness of information? Why? If not, what is missing?

In addition, the content needs to be evaluated in terms of correctness, accuracy and its appropriateness for intended users. The content needs to be presented in a way that is useful and understandable by the intended audience (i.e., are definitions of terminology likely to be unfamiliar to the audience?). Users and content experts may have different opinions about this.

Results from the evaluation of component 2 are reported to the project team, with recommendations regarding changes to be made to web design and web content.

## **2.3 Evaluation of use of the educational resource (Component 3)**

In order to gather data on the use of the resource, it is helpful to survey users of the web resource and also non-users. Data on use can be collected by internal electronic tracking to determine how many people are using the resource and how frequently they are using the site. Online surveys can also be given to users, asking them to classify themselves in terms of how frequently they use the website. People who have used the site once or a few times only can be asked why they have not used it more.

It is also important to determine which resources are used and why, as well as which resources are not used, and why not. This will provide information about which components of the resource need improvement. Questions to ask users include:

1. Are they aware of the different areas that are available on the website?
2. How useful are the different areas of the website? Why?
3. Why do they use certain areas and why don't they use other areas?
4. What do they like best about the website and why?
5. What do they like least about the website and why?
6. If and how will they keep using the resource. If not, why?
7. Which parts of the website motivates them and why and which parts discourages them and why?

It can be informative to learn whether lack of use is related to web design. This information can be used to determine areas for improvement to be made to the web design. Questions to use include:

1. Can intended users not access the resource?
2. Can intended users not access the resource in ways that are acceptable to them?
3. Is the website not easy to navigate?
4. Is the website not aesthetically appealing?
5. Are people not able to find content that they are looking for and that is available on the resource?
6. Do people not trust the resource?

A related question is whether lack of use is related to web content. This information can be used to determine areas for improvement to be made to the web content. Here, a different set of questions may be used, such as:

1. Is the content not relevant to them? If so, why not?
2. Is the content not covering the needed scope? If so, where not?
3. Is the content not valid? If so, why?
4. Is the content not complete? If so, where are the gaps?
5. Is the content not accurate? If so, how?
6. Is the content not appropriate for them? If so, why?

At the end of this third component, there is a report of evaluation findings with recommendations of changes to be made to web design and web content that will increase the use by overcoming problems identified by users and non-users.

## 2.4 Evaluation of Educational Impact (Component 4)

While it may seem intuitive to evaluate the quality of web content and design, many evaluations of online educational resources fail to look beyond these characteristics to the actual impact of the resources. To gather data on educational impact different types of data may be gathered. There are several ways to gather data on impact. One way is to

examine the impact on instructors' perceptions of the content area and the second is to determine impact on the educational environment. An important question to examine is whether the use of this resource by educators has a positive impact on classroom instruction, on curriculum, and/or on assessment practices.

The impact on students also needs to be considered. This can be done by administering students' surveys on affective variables (attitude and motivation) as well as tests measuring student learning (content knowledge). In addition, it is important to determine whether the resource has a positive impact on students' perceptions of the content domain.

The sustainability of the impact of the resource also needs to be examined. The goal of an educational website is usually to have a long-term impact so a longitudinal study may be conducted. It may also be useful to determine how well-known the resource is in the professional community; by finding out if other websites link to this one, if the resource is mentioned at professional conferences, in publications, and/or in textbooks.

Again, a report of evaluation is given to project leaders at the end of this phase, and recommendations are offered regarding changes to the web design and content to improve impact.

Even though the evaluation model is divided into four components one should know that these components interact and are influencing each other. The combination of the web design component and the web content component determines the educational value of the resource. The educational value determines the use of the educational resource, which in turn determines educational impact.

### 3. USE OF THE IEM TO EVALUATE THE ARTIST PROJECT

We now illustrate how the IEM was used to evaluate an online educational resource: the ARTIST website. NSF funded the Assessment Resource Tools for Improving Statistical Thinking (ARTIST) project (DUE-0206571) to address many current assessment challenges 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/>) 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, in evaluating individual student achievement, in evaluating and improving their courses, and in assessing the impact of reform-based instructional methods on important learning outcomes.

The online resource includes the following products:

- A collection of over 1,000 expert-reviewed 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 rich text format (rtf) 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.).
- Separate online tests that measure conceptual understanding in 11 important areas of a first course in statistics which have high validity and reliability.
- The Comprehensive Assessment of Outcomes in a first Statistics course (CAOS), a test that measures statistical literacy and reasoning (see delMas et al., 2007).

The IEM was used to plan the ARTIST evaluation, to design instruments and collect data, to determine how instructors were using the materials and how the materials were helping to achieve the original project goals related to improving the teaching and learning of statistics. The ARTIST Project Directors were involved in the planning and decision making of every step of the evaluation. The evaluation took place over the course of one year.

Six instruments were developed and administered during the evaluation. Those were a) a pre evaluation interview protocol, (b) a survey for ARTIST users, (c) a survey for non-ARTIST users, (d) an interview protocol for frequent ARTIST users, (e) an observation protocol for frequent Assessment Builder users, and (f) an observation protocol for first time Assessment Builder users.

Table 1 provides an overview of the evaluation process, methods, and instruments used to collect data during each of the components of the evaluation model. These are arranged according to the four components or phases of the IEM. It is important to point out that the evaluation instruments were not all created at the same time. The survey for ARTIST users was first created and administered, and then the data analyzed in order to tailor the next data collection instrument for the users, namely the interview protocol for frequent ARTIST users, to the evaluation needs. The evaluator also created the observation protocols for the frequent and the first time users of the Assessment Builder only after analyzing the data from the ARTIST users' survey.

Table 1

*Overview of Evaluation Process, Methods, and Evaluation Instruments per Component of the IEM*

Components of IEM	Method
Component 1: Evaluation Planning	Meetings to plan evaluation Pre-evaluation interviews ARTIST users and non-ARTIST users
Component 2a: Evaluation of web design	Survey for ARTIST users Survey for non-ARTIST users Observations of frequent Assessment Builder users Observation protocol for first time Assessment Builder users Reporting on web design
Component 2b: Evaluation of web content	Survey for ARTIST users  Survey for non-ARTIST users Reporting on web content
Component 3: Evaluation of use of the educational resource	Survey for ARTIST users Interviews of frequent ARTIST users Reporting on use of educational resource
Component 4: Evaluation of educational impact	Interviews of frequent ARTIST users Reporting on educational impact

We now provide details on the various instruments developed and used in the ARTIST evaluation that emerged from using the IEM for each component of the evaluation.

#### 4.1 A Pre-Evaluation Interview Protocol ARTIST Users and non-ARTIST Users.

The pre-evaluation interview protocol for ARTIST users and non-ARTIST users was created and administered to conduct a needs assessment. The interviews were conducted to collect information about statistics instructors' assessment practices. This information was used to inform the development of the other evaluation instruments. Invitations for participating in the interview were sent by e-mail and 9 interviews were conducted, lasting between 30 and 70 minutes. Some of the results from these interviews include that statistics educators, even those with many years of teaching experience, find it

challenging to write good assessment items. Most interviewees used other resources to find assessment items but found it time-consuming to find items of good quality that would meet their standards. Users of the ARTIST website commented positively about this resource in terms of quality and usability. It became apparent that non-users had a misconception about the goal of the ARTIST website and thought the website was developed to be used by students rather than educators.

## 4.2 Online Survey of ARTIST users

An online survey of ARTIST users was developed and administered to collect information about the web design, the web content, and implementations in educational settings. The survey contained 50 questions. The survey was pilot tested with 6 people. All registered ARTIST users teaching introductory statistics in the US, who had accessed the website since August 2004, and who were teaching that semester, received an e-mail invitation for submitting the online survey. The e-mail contained a link to the survey and the first 100 respondents were promised a gift certificate as an incentive. Ninety-eight statistics educators responded to the user survey which provides a rich set of data to reveal how the intended audience was actually using the web resources. Some of the results indicated that more than twice and more than half of the respondents had visited the website more than 5 times. In terms of web content, the respondents were quite positive and did not indicate the need for major changes. The website appeared to be reliable and links labeled in a way that was descriptive of their content. The organizational scheme received high ratings as well. Suggestions were made regarding the design of the Assessment Builder as some users found some tasks challenging. Not all pages of the website seemed to be used as frequently, however, of those who had browsed the pages the majority found them useful. In general, users found the ARTIST assessment items to be of the same or higher quality than items they selected from other resources. The ARTIST items were judged to focus more on conceptual understanding and less on computation, which was what the project intended.

## 4.3 Survey for Non-ARTIST Users

A unique aspect of this evaluation was to attempt to survey intended users who had not used the ARTIST resources. The survey for non-ARTIST users was created and administered to collect information about characteristics of statistics instructors who had not chosen to utilize ARTIST resources. The survey contained 19 questions. This survey was not pilot tested since it contained the same questions as the ARTIST users' survey, with the exception of 2 additional questions. Invitations for submitting the survey were sent by e-mail to several list serves that are used by statistics instructors (e.g., Isostat group, ASA Stat Ed, and the SIGMAA listserv). The e-mails contained a link to the survey. Respondents had the possibility to submit their e-mail address and 3 \$20 gift certificates were given to a random selection of 3 respondents. Eighty-nine statistics educators responded to the survey and provided information on why they had not used the ARTIST materials. The main reason for non-users not having used the website was that they were not aware of the website or had misconceptions about its goals and thought the website was developed to be exclusively for online testing rather than as a resource to

statistics teachers. For many of the survey respondents, the survey actually triggered them to discover and explore the ARTIST website for the first time, leading them to eventually utilize the resources there.

#### 4.4 Interview Protocol for Frequent ARTIST Users

The interview protocol for frequent ARTIST users was used to collect more detailed information about use of ARTIST resources in different settings as well as information on the perceived educational impact of this use. Participants were selected based on the number of ARTIST topic scales and ARTIST items they had administered to their students. Invitations for participating in the interview were sent by e-mail. Volunteers were given a gift certificate. Seven interviews were conducted, lasting between 20 and 30 minutes. Many valuable types of information were obtained from the interviews, including positive reviews of the ARTIST website and the assessment items in particular. Users found the assessment items to be helpful in revealing valuable information about their students' understanding of statistics, as well as students' lack of understanding of important statistics ideas. This information led the instructors interviewed to reflect upon and suggest changes to their teaching. While at the time of the interviews, none of the users had implemented major changes to their teaching practices, they were planning on doing so for the next academic year.

#### 4.5 Observation Protocol for Frequent Users of the Assessment Builder

An observation protocol was developed to study people who frequently used the ARTIST site build tests using the item data base, named the Assessment Builder. The information collected related to the web design and the difficulty level of navigation and use. We wanted to learn where the Assessment Builder was difficult to use, even for experienced users. Invitations for participating in the observation study were sent by e-mail and volunteers were given a gift certificate. Three observations were conducted. Participants were asked to conduct a think-aloud while performing certain tasks, such as create a test, add a question to that test, remove a question. The observations of frequent users produced many practical suggestions for improvement. For example the frequent users did not want to automatically receive the same instructions about how to use this tool but wanted it to be available in case they wanted to revisit some components of the instructions. Frequent users also voiced a request for a search engine to be added that would display newly added assessment items only. There was also a request for some of the graphical items to be improved in terms of visual quality. All of these suggestions came about by commenting on the use of the website while actually accessing and using the resources, and were used to improve the website to better meet the needs of the frequent users.

#### 4.6 Observation Protocol for First-Time Assessment Builder Users

It was also important to observe first-time users of the ARTIST site and Assessment Builder. A protocol was designed to collect information about the web design and to learn where the Assessment Builder was difficult to use for first-time users. Invitations

for participating in the observation study were sent by e-mail and volunteers were given a gift certificate. Two observations were conducted, lasting between 30 minutes to one hour. Participants were asked to conduct a think-aloud while performing certain tasks, such as create a test, add a question to that test, remove a question. These observations provided valuable information about the more challenging tasks that are part of the test-building process. For example, first-time users asked for the login page to be changed and certain navigation buttons to be renamed. They also would prefer the option of all selected items to be displayed on one page instead of on multiple pages. Again, these suggestions were used to make changes in the website and its use.

## 5. VALUE OF THE EVALUATION

The data gathered provided both formative and summative information about ARTIST. Formative information revealed information about the web design and the web content, and included recommendations for improvement of the design and the content of the ARTIST materials and website. Formative information also concerned the use of the educational resources. We learned how users use the resource and why non-users did not use the resource. This provided information about how to motivate non-users to start using the online educational resource.

Summative information concerned the use of the educational resource and its educational impact. We learned that ARTIST is a valued resource and used by many introductory to statistics educators. We also learned that the existence of ARTIST has had an impact on users' assessment instruments and assessment strategies, but that it has not impacted instructional methods and therefore had not impacted student outcomes yet at the time of the evaluation. There was a gap in how the IEM was implemented in this case study. The standards, goals, and criteria for quantifying the quality of the resource had not been set during the evaluation planning of the ARTIST project. The lack of these standards and criteria made it impossible for the evaluator to make conclusions about the level to which the goals of the projects had been achieved. These results imply that standards, goals, and criteria for quantifying the quality of the resource need to be established during the evaluation planning.

Evaluation reports were created based on all the data collected by the evaluation instruments. During a meeting with all the ARTIST directors, the evaluator reported on the findings. The ARTIST PIs received a copy of the summary of the findings, a copy of all instruments, and all the raw data after identifiers had been removed. The evaluation findings were reported in terms of web content, web design, use of the educational resource, and educational impact.

Results were summarized and examined for the four components of the evaluation model. It is important to note that implementing Component 4 of the evaluation model only makes sense if the web resource has existed for some time. Newly created online educational resources can not have an educational impact if instructors have not used the resource for some time. It takes time for instructors to take action based on the newly developed ideas. Therefore, some educational resources may not be ready to be evaluated

in terms of educational impact. Based on this lesson learned, the author added to the model that a standard has to be set by the educational resource directors in terms of how soon they expect the program to have an educational impact. This standard can then be used when interpreting data.

Another important note is the use of incentives to encourage people to respond to surveys and participate in interview and observations. A supplement to the ARTIST budget included stipends to use as incentives. We think that the incentives were important and resulted in valuable data by compensating people for their time. We suggest that other statistics education projects include a budget item or incentives as part of their evaluation.

The value of this evaluation model was demonstrated by the types of information gathered via all the different methods and tools and how these were used to improve the website. What is unique about this model is that it is tailored to the evaluation of online educational resources where other evaluation models and evaluation approaches are more general. For example, having the evaluator survey both users and non users of the website, and observing both new and frequent users, would not necessarily be part of an evaluation designed using a generic evaluation model.

## 6. SUMMARY

Although professional evaluators are aware of different models to use in evaluating programs, curriculum, or courses, statistics educators may not be aware of the use of a particular model to guide an evaluation. The IEM which was developed by and used to evaluate all aspects of the ARTIST project may be utilized in evaluating other projects as well. The different components of the model suggest areas to examine and types of questions to use, and look beyond the more apparent features of web design and web content to consider how the materials are used and the impact they may be having on teachers and students.

The use of the model revealed many important findings about how people perceived and used ARTIST materials, and suggested changes to improve the website and its resources. We recommend the use of this model for other online instructional resources in statistics education, as well as in other disciplines. For example, the IEM has been used to evaluate the U.S. National Institutes of Health' National Diabetes Education Program web site (<http://www.ndep.nih.gov/>) and the U.S. National Institutes of Health has expressed an interest in applying the model to evaluate additional online resources. We encourage professionals who are developing or revising websites with educational resources to consider utilizing this model in order to gather useful data to improve their online resources.

## REFERENCES

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999), "*Standards for Educational and Psychological Testing*". Washington, DC: American Educational Research Association.
- Branch, R.M., Kim, D., and Koenecke, L. (1999), "*Evaluating online educational materials for use in instruction.*" ERIC Digest. ERIC Clearinghouse on Information and Technology, Syracuse, NY. ED430564.
- Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007, October). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education Journal*, 1(1). (<http://repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art2/>)
- delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28-58. [http://www.stat.auckland.ac.nz/~iase/serj/SERJ6\(2\)\\_delMas.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ6(2)_delMas.pdf)
- Fitzgerald, M.A., Lovin, V., and Branch, R.M. (2003), "*The gateway to educational materials: An evaluation of an online resource for teachers and an exploration of user behavior.*" *Journal of Technology and Teacher Education*, 11(1), 21-51.
- Fitzpatrick, J., Sanders, J., & Worthen, B. (2004). *Program Evaluation: Alternative Approaches and Practical Guidelines* (3rd ed.). Boston: Allyn & Bacon.
- Garfield, J. & Gal, I. (1999). Assessment and Statistics Education: Current Challenges and Directions. *International Statistical Review*, 67, 1-12.
- Jenkins, A.G., and Robin, B. (2002, April), "*Evaluation of an educational website for the Bayou Bend collection and gardens, museum of fine arts, Houston.*" Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Kane, M.T. (2001), "*Current concerns in validity theory.*" *Journal of Educational Measurement*, 38(4), 319-342.
- Kleiner, A., and Lewis, L. (2003), "*Internet Access in U.S. public schools and classrooms: 1994-2002* (No. NCEES 2004-011)." Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Marchionini, G. (2000), "*Evaluating digital libraries: A longitudinal and multifaceted View.*" *Library Trends*, 49(2), 304.
- Messick, S. (1989), "*Validity*". In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp.12-104). New York: American Council on Education and Macmillan.
- Moss, P.A. (1995), "*Themes and variations in validity theory.*" *Educational Measurement: Issues and Practice*, 14(2), 5-13.
- Ooms, A. (2005), "*The Iterative Evaluation Model for Improving Online Educational Resources.*" Doctoral Thesis, University of Minnesota.
- Sax, G. (1997), "*Principles of educational and psychological measurement and evaluation* (4th ed)." Belmont, CA: Wadsworth Publishing Company.
- Smith, E.R., & Tyler, R.W. (1942), *Appraising and Recording Student Progress*. New York: Harper & Row.

Stufflebeam, D. (2000) The CIPP model for program evaluation. In D. Stufflebeam, G. Madaus, & T. Kellaghan, *Evaluation Models: View-points on Educational and Human Service Evaluation*. Boston: Kluwer-Nijhoff, pp. 279-315.

Stufflebeam, D. L. and Shinkfield, A J. (2007); *Evaluation Theory, Models, and Applications*, San Francisco: Jossey-Bass.

Web-based Education Commission. (2000), "*The power of the Internet for learning: Moving from promise to practice: Final report of the Web-Based Education Commission.*" Washington, DC: The Commission. Retrieved November 12, 2004 from <http://www.ed.gov/offices/AC/WBEC/FinalReport/WBECReport.pdf>.