Teacher's Corner

Six Online Statistics Courses: Examination and Review

Jorge LARREAMENDY-JOERNS, Gaea LEINHARDT, and Javier CORREDOR

We extend George W. Cobb's evaluative framework of statistics textbooks to six online instructional materials that exemplify the diversity of introductory statistics materials available on the Internet. Materials range from course Web sites with limited interactive capabilities to courseware and electronic textbooks that make extensive use of interactive learning objects and environments. Instructional materials are examined in light of recent cognitive research that underscores the robustness of learning from examples, the importance of authentic problem solving in promoting knowledge in use and skill acquisition, and the use of feedback to maximize learning opportunities. Selected units that focus on statistical tools (measures of central tendency, simple linear regression, and one-way analysis of variance) are analyzed in terms of authenticity and diversity of examples, authenticity and cognitive complexity of exercises, and use of interactive learning objects and feedback. General conclusions and suggestions for future directions for online statistics instruction are presented.

KEY WORDS: Distance learning; Evaluation; Web-based instruction.

1. INTRODUCTION

In 1987, George W. Cobb examined 16 introductory statistics textbooks in the *Journal of the American Statistical Association*. Cobb laid out an evaluative framework that considered technical level and quality of exposition, topics covered, and quality of exercises. He selected four standard topics: sample mean, sample standard deviation, normal distribution, and sampling distribution of the mean. He characterized explanations by identifying the extent to which the expositions relied on formulas and derivations. Cobb estimated the breadth and depth of explanations by comparing the content covered and the level of detail within an additional set of topics (regression, analysis of variance, exploratory data analysis, and computers). Finally, he addressed the quality of exercises by estimating the authenticity of datasets, the meaningfulness of the studies described in the problem statements, and the ratio of thinking to number crunching.

In the 17 years since Cobb's evaluative framework was published two major changes have occurred in the landscape of statistics education: First, there has been growing recognition of statistical knowledge as a crucial component of core scientific literacy (Utts 2003). As a result, we now see the teaching and learning of statistics in elementary, secondary, and higher education (NCTM 2000). Second, there has been a flowering of online technologies and courses that both support and teach statistics. The use of online technologies is often predicated under the assumptions that the Internet can contribute to making statistical knowledge accessible to vast audiences and that online multimedia environments can make learning more interactive and meaningful. Today an impressive variety of instructional materials in statistics education is accessible on the Internet, from full stand-alone introductory courses and electronic textbooks to digital repositories of learning objects and datasets.

This article extends Cobb's evaluative framework to a set of online instructional materials. Like Cobb, we focus on the quality of explanations and exercises. Yet, in identifying their critical features we draw on recent cognitive research that underscores the prevalence of learning from examples, the importance of genuine problem solving in promoting knowledge in use, and the use of feedback to maximize learning opportunities. Our goals are to provide a sense of the quality of some instructional materials available on the Internet and to suggest criteria for inspecting such materials; criteria that we imagine will be expanded and developed over time as more materials and resources emerge.

2. EVALUATIVE GOALS AND CRITERIA FOR ONLINE COURSEWARE

An important goal for the evaluation of online multimedia courseware is the assessment of instructional explanations and learning opportunities. In order to assess explanations and learning opportunities three analyses need to be conducted. First, conceptions of learning and teaching that underlie the design of instructional materials must be explicated. Second, the extent to which instructional materials comply with well-established principles of learning and teaching should be assessed. Third, learning affordances and constraints linked to the technical implementation of courseware need to be specified. [We use the term "affordances" following Gibson's (1977) sense of the term as it has been adopted by psychologists and others (Greeno 1994). An affordance is an opportunity in an environment to make use of a physical or mental resource to accomplish a goal. But the op-

Jorge Larreamendy-Joerns is Associate Professor, Department of Psychology, Universidad de los Andes, Bogota, Colombia. Gaea Leinhardt is Professor, and Javier Corredor is Doctoral Student, University of Pittsburgh, PA (E-mail: gaea@pitt.edu). Support for this article was provided by a grant to the second author from Carnegie Mellon University via The William and Flora Hewlett Foundation. The opinions expressed do not necessarily reflect the position or the policy of the foundation and no official endorsement should be inferred.

 DOI: 10.1198/000313005X54162

portunity only exists if one recognizes it as such. The computer environment offers instructors and designers affordances in the area of dynamic relationships, for example, but teachers who have been constrained for more than 30 years to static representations of important ideas may not recognize these opportunities and, consequently, may not make use of them.] Doing this type of analysis documents whether online environments make effective use of multimedia and Internet possibilities in light of given educational goals. In this review we focus on: (a) examples as core components of instructional explanations; (b) exercises or problem-solving venues as environments that support and foster learning by doing and skill acquisition; and (c) interactive learning environments as distinctive features of online courseware.

2.1 Examples as Core Components of Instructional Explanations

Cognitive research suggests that examples are a powerful tool for learning and instruction (Anderson, Farrell, and Sauers 1984; Fergusson-Hessler and DeJong 1990; Lavigner and Lajoie 1995; McCarthy 2002; Mitchell, Keller, and Kedar-Cabelli 1986; Rissland 1989; Simon and Zhu 1988). Examples provide a basis for inductive generalization, help students fix the reference of abstract concepts, allow students to connect new information with prior knowledge, and constitute essential criteria for indexing and organizing knowledge in memory (Anderson 1993; Rissland 1989). Studies show that, when learning from textbooks, students (especially good ones) spend considerable time studying worked out examples and tend to rely heavily on them, even at the expense of explicit instructions and explanations (Anderson, Greeno, Kine, and Neves 1981; LeFevre and Dixon 1986; McCarthy 2002). However, for examples to be effective in supporting learning, some conditions must be satisfied. First, there needs to be more than one example. Students learn more when multiple instances that uniformly illustrate critical features of problems allow them to discriminate between incidental and relevant features (Quillici and Mayer 1996).

Second, examples need to be authentic in the sense of being connected to the understanding of students and realistic within a domain (Leinhardt 2001). Authenticity concerns the extent to which problem statements or example situations involve statistical practices that mirror those of statistics in use and that are connected to meaningful queries within statistics or science. This is what Shaffer and Resnick (1999) called "methods of inquiry aligned with the discipline." Paradigmatic cases of authentic, yet instructionally tractable examples are Fisher's classical analysis of species of irises for taxonomic purposes (Fisher 1936) and Tukey's formal questions about data display and reduction (Tukey 1977). Even though authentic examples and tasks may contain information that is close to the learner, authenticity is not the same as familiarity. Authenticity does not imply a commitment to a purely pragmatic view of statistics at the expense of an appreciation of mathematical properties.

Finally, examples need to vary among each other in significant ways. Although prototypical examples allow students to quickly develop a problem schema, variability exposes students to atypical situations, thus increasing transfer of learning (McCarthy 2002; Pirolli 1991). In statistics, examples can vary in terms of the distributional properties of their datasets. For example, different instances of simple regression can involve data with and without extreme scores, or analysis of variance can be illustrated for balanced and unbalanced designs. Examples can also vary in terms of their content (e.g., sociology, economics) or their research question and design within a content domain (cover story). Variability in the content domain or cover story of examples provides students with opportunities to connect learning materials with their understandings, interests, and prior knowledge.

2.2 Exercises as Opportunities for Skill Acquisition and Knowledge in Use

Problem-solving environments support practice, which causes continuing improvement in speed and accuracy of performance. Further, if problem solving is complemented with feedback and learners engage in generating alternatives, causal attribution, and hindsight, tasks may become an occasion for learning by doing (Langley and Simon 1981). Exercises also allow important connections to "knowledge in use" to be built up, thus reducing the potential for inert knowledge to float detached and unusable (Leinhardt 2001). From a disciplinary perspective, exercises play a role analogous to that of examples in that they are informative of the kinds of problems that define a disciplinary domain.

However, the benefit that students can derive from exercises depends greatly on their frequency, cognitive complexity, and authenticity. It depends on their frequency because the psychological processes and knowledge structures that underlie skilled problem solving (e.g., general heuristics, domain-specific strategies, rule-based systems, problem schemas) are consolidated only after considerable practice with the problems typical of a domain (Anderson 1993; Dunbar 1998). The students' benefit also depends on the cognitive complexity of the exercises. By cognitive complexity we mean the extent to which the task requires multiple inferential steps, involves multiple possible solutions, and allows for multiple paths to a solution (Stein and Lane 1996; Sweller and Chandler 1994). Complex tasks press students to consider a space of plausible solution paths, to evaluate solution strategies, and to activate knowledge required to draw relevant inferences and make decisions. In contrast, tasks that rely exclusively on rote memory or execution of algorithms may mask the students' lack of deep understanding of the meaning of concepts and the conditions under which procedures should be executed (Lovett 2001). Finally, the students' benefit from exercises depends on the degree of authenticity of problem-solving tasks. Authentic tasks—that is, tasks that require students to engage in intellectual moves that mirror those of professionals in the discipline or that resemble practices that will be meaningful to the students in their everyday life—foster transfer of learning and knowledge in use.

2.3 Interactive Learning Environments as Distinctive Features of Online Courseware

One promise of online education is the increase in the quality and frequency of student-content interaction. Interactivity gives students the opportunity to act upon problem states and representations, and provides them with information about the effects of their actions and the state of their understanding. The quality of the interaction is a function of the nature of the actions that learners are allowed to carry out and the quality and timing of the feedback provided. Actions can be as restricted as setting in motion a predetermined process, as when students simply click on an applet that demonstrates a fixed procedure. Actions can also be flexible as when students select their own problem-solving strategies in the context of a virtual lab or an intelligent tutor, or as when they set at will the parameters of a simulation.

In the current state of development of computing and online technologies, interactivity is most ubiquitous at the level of representations (i.e., animated diagrams) and feedback about knowledge assessment questions. Representations are a supporting pedagogical mechanism in explanations (Leinhardt 2001). Graphs, tables, and diagrams convey visually complex information, are a hallmark of disciplines, and help learners focus on relations rather than isolated pieces of information (Funkhouser 1937; Kosslyn 1989; Larkin and Simon 1987; Leinhardt, Zaslavsky, and Stein 1990; Shah and Hoeffner 2002). Online representations (e.g., histograms, regression lines) can be dynamic in that they can be operated upon and coordinated with other representations, and in that they change over time to visualize processes. Additionally, since representations are easily generated, instruction can focus on the connections between representations and concepts, and not on the nuts and bolts of graphing (Wender and Muehlboek 2003).

However, although applets and interactive learning objects are powerful learning tools, their use is not without concerns. First, learning through interactive objects depends on performing the appropriate actions and noticing critical features. Second, interactive learning objects may include representations that are transparent to the expert eye, yet intractable to students. Third, it is important to understand that interactive learning objects support explanations, they do not replace them. No amount of visualization can compensate for a thoughtful explication of concepts, processes, and procedures against the backdrop of principles.

Feedback is a critical component of highly effective learning environments, such as one-to-one human tutoring, and unfortunately one that is often rare in traditional classrooms. The effectiveness of feedback is a complex function of its timing (i.e., immediate or delayed), its informativeness (i.e., the amount and quality of information it provides to the learner about performance and understanding), and the contingencies of its administration (i.e., when learners fail to perform, when they succeed, or both) (Kulik and Kulik 1988; McKendree 1990).

3. METHODOLOGY

3.1 Selection of Online Courses

We conducted an Internet search of online statistical courseware, using four criteria for inclusion: first, the courses needed to be aimed at the introductory, noncalculus based level; second, the courses had to be stand-alone courses that did not require necessarily the intervention of an instructor; third, the courses had to be available to distance learners, not only to enrolled campus students; and fourth, there had to be a complete version available. We also wanted to select a sample of courseware that might exemplify the diversity of products currently available on the Internet, from courses that are essentially translations to HTML of otherwise printed materials to highly interactive courseware, and from courses that are the result of collegial teamwork and are disseminated through commercial venues to courses that have grown out of a single author's interest in the scholarship of teaching. We considered courseware advertised as online courses, course Web sites, or online textbooks. [Since we wanted to explore courseware delivered over the Internet because of its potential to reach vast audiences and to provide adaptive feedback, some very good courses were excluded, for example, Paul Velleman's ActivStats, which makes effective and extensive use of videos to contextualize examples and voice-over animations to support instructional explanations.] The following six online materials were selected:

1. CyberStats (http://statistics.cyberk.com/splash/)
2. StatCenter Psych 3000 Online (http://www.psych.utah.edu/stat/)
3. Introductory Statistics: Concepts, models, and applications (http://www.psychstat.smsu.edu/sbk00.htm)
4. Investigating Statistics (http://espse.ed.psu.edu/statistics)
5. Seeing Statistics (http://www.seeingstatistics.com)
6. SurfStat (http://www.anu.edu.au/nceph/surfstat/surfstat-home/surfstat.html)

3.2 Course Descriptions

3.2.1 CyberStats

CyberStats was developed by Alexander Kugushev, authored by a team of 21 faculty members at 18 different colleges and universities, and delivered online by CyberGnostics, Inc. CyberStats can be used by institutional instructors (for which a course management system is provided) or by self-learners. The registration fee for an individual user is below $50. CyberStats includes more than 100 interactive applets, practice materials with feedback capabilities, integrated statistical software, knowledge assessment and testing facilities, nearly 200 datasets, and a set of online tools (from calculators and search engines to note taking facilities). The course consists of 42 units grouped into seven major topics: collecting and visualizing data, modeling random behavior, inference, regression, design of experiments and ANOVA, time series, and statistical process control. Instructors can customize their own syllabus by selecting the number and sequence of instructional units. All units include a set of introductory problems, exposition of core terms and concepts, application of concepts to practical problems, constraints on and misuses of procedures, and worked-out integrative examples. Practice problems are located throughout each unit and additional exercises are given at the end.

3.2.2 Psychology 3000 Online

Psychology 3000 Online (hereafter Psych3000) is an online course developed at the University of Utah. Psych3000 is offered, on an academic term basis, to University of Utah regular students and to independent learners, through continuing education, for a cost of about $600. Psych3000 is a suite of instructional components supported by an Oracle database, a set of Java applets, and HTML Web pages. The course includes printable lectures, downloadable PDF files for note taking and practice homework, applets, an integrated statistical software, online statistical tables, online practice facilities, instructional games, chat and electronic discussion facilities, and a virtual lab. Psych3000's virtual lab allows students to research questions set by the instructor in a simulated reality. Students review literature, select variables, collect data, perform statistical analyses, and write a final report bearing on the initial research questions. Psych3000 includes 23 thematic units that cover basic probability, science and statistics, statistical distributions, measures of center and spread, correlation, regression, and hypothesis testing.

3.2.3 Seeing Statistics

Seeing Statistics is a Web-book authored by Gary McClelland of University of Colorado at Boulder and published by Duxbury Press. The paperback edition (which provides access to the online version) currently costs under $50. Through the use of approximately 70 instructional applets, Seeing Statistics emphasizes the role of visualization and interactivity in the learning of statistical concepts. Each Web page in Seeing Statistics includes links to a set of tools (contents, calculator, glossary, search engine, references, feedback to designers, and site help), and to supporting questions and additional examples and explanations. Seeing Statistics can be used as a traditional textbook, as a source of dynamic graphs suitable for classroom use, as a complement for lab activities, and as a venue for distance learning. Contents cover graphical displays, measures of center and spread, probability, normal distribution, inference and confidence, regression and correlation, and one- and two-sample comparisons (analysis of variance is not covered in this version). Practice exercises are provided at the end of each thematic unit.

3.2.4 SurfStat

SurfStat is an online course developed by Keith Dear, Rod Smith, Jonathon Coombes, and Robert Brennan at the University of Newcastle, Australia, and presently hosted at Australian National University. SurfStat is the evolution of an existing set of course notes (translated to HTML to implement hypertext links) and added Java applets to replace statistical tables and illustrate concepts. Access to SurfStat is free, although it may be purchased for under $200 and mounted on a local Web to increase processing speed. Every Web page in SurfStat has a control panel on the left of the screen with hyperlinks to home, glossary, tables, feedback, help, and product information. Contents are organized in five major units: summarizing and presenting data, producing data, variation and probability, statistical inference, and control charts. Some units have "progress checks" (i.e., multiple-choice items with feedback).

3.2.5 Investigating Statistics

Investigating Statistics is a course Web site developed by Robert Hale of Penn State University and supported by Wadsworth Corporation. Investigating Statistics is currently used as a learning resource for statistical courses in educational psychology at Penn State. Investigating Statistics includes a search engine, a series of applets for statistical analysis and graphics, QuickTime movies, and hyperlinks to several statistical sites on the Internet. Computer-scored tests can be taken at the end of each chapter. Navigation through Investigating Statistics is simple because each thematic unit consists of a single Web page, with a navigation frame to the left of the screen. Contents are grouped into 16 chapters that cover scientific method and data analysis, exploratory data analysis, measures of central tendency and spread, probability, inferential statistics and hypothesis testing, sampling distributions, two-sample tests, ANOVA, regression, nonparametric analyses, and ethics.

3.2.6 Introductory Statistics: Concepts, Models, and Applications

Introductory Statistics (hereafter Intro Stats) is an electronic textbook authored by David Stockburger, published and Web-supported by Atomic Dog publishing, and sold for under $50. Intro Stats includes applets, search capabilities, calculator, glossary, table of contents, interactive end-of-chapter quizzes, summaries and key terms, and highlighting and note taking facilities. Instructors can add their own online text notes, hyperlinks, instructions, and knowledge assessment items. This Web-book stresses the notion of model and the understanding of mathematical definitions of statistical terms rather than inductive learning from examples. Intro Stats consists of 23 chapters that cover scientific and mathematical models, fundamentals of algebra, frequency distributions, models of distribution, transformations, regression and correlation, hypothesis testing, and probability.

3.3 Coding and Analysis

To examine the online materials, four topics were selected based on their disciplinary importance and conceptual difficulty for beginning learners: measures of central tendency, simple linear regression, one-way analysis of variance, and sampling distributions. Units or course segments corresponding to the first three topics were analyzed in terms of the quality of examples, the quality of exercises, and the nature of online interactivity. Units or course segments on sampling distributions were reserved as a special case for the analysis of online interactivity.

3.3.1 Analysis of Examples

For coding purposes, we defined an example as a single dataset or problem scenario used to illustrate a target concept or procedure. We did not recount core examples that were used subsequently as carryover illustrations. We also excluded passing examples, that is, examples that were not unpacked and that did not exceed a verbal clause or sentence.

Authenticity of examples. To evaluate the authenticity of examples, we assessed the degree to which the example provided a meaningful description of the research question that originated the data, the nature of the variables involved, and other relevant information required to make sense of statistical analyses. We adapted Cobb's (1987) three-point scale for authenticity and assigned each example to one of the following categories: schematic (rank 0), cursory (rank 1), and thorough (rank 2). Schematic examples consist of datasets with no cover story or with a description so broad that the identity of variables could be changed without any effect on the pattern of analysis and inferences. Cursory examples include contexts or research settings that are plausible, yet cryptically described, and that could be greatly enhanced by additional information. Finally, thorough examples provide a reasonable description of the setting in a way that makes the example engaging and fully understandable. Following Cobb (1987), we computed for each course the average rank across units, and multiplied the result by 50 (to have a scale from 0 to 100). Twenty-five percent of the examples was coded for reliability by two independent coders. Overall intercoder reliability was 96.7%.

Variability of examples. We developed two approximate measures of example variability. The first was the average number of examples, across thematic units, from different content domains (e.g., psychology, economics) or example cover stories within a single domain. Examples consisting only of numerical datasets, with no cover story, were excluded from this counting. This measure gave us a sense of the extent to which the courses made an effort to provide examples illustrating uses of statistics in different contexts of inquiry. Yet, because superficially different examples can be conceptually equivalent (in that they illustrate the same aspects of a concept or a procedure), we devised a second measure to estimate whether examples were conceptually distinct; that is, whether different examples were explicitly crafted to illustrate different aspects of concepts or procedures (or special cases of them). For each unit, we indexed the concepts illustrated in each example, identifying both ordinate (i.e., simple linear regression) and subordinate (i.e., simple linear regression with outliers) concepts. We counted the number of examples that were conceptually different at a subordinate level and averaged the number of conceptually distinct examples across units for each course.

3.3.2 Analysis of Exercises

In each course, we identified exercises dispersed through the thematic units, located at the end of each unit, or placed at a given location in the course Web site. We counted as a single exercise series of questions referring to a single problem statement or dataset. Then we evaluated each exercise in terms of its authenticity (using the same three-point scale used for the examples) and its cognitive complexity. To assess complexity, we adapted a coding schema developed by Stein and Lane (1996), which underscores the idea that cognitive complexity increases when students need to make sense of tasks by themselves, defining strategies, making decisions, interpreting results, and drawing conclusions (i.e., when tasks have greater degrees of freedom). Complex tasks differ from tasks where students only have to execute algorithms or recall information from memory. We added a category for tasks where there are no performance standards and where cognitive difficulty depends not only on the task itself but also on the student's level of engagement. We assigned each of the exercises to one of the following categories (ordered in decreasing degree of complexity): complex use of procedures, simple use of procedures, direct recall, and no performance standards (category definitions are given in Table 1). Twenty-five percent of the exercises was coded for reliability by two independent coders. Overall intercoder reliability was 94%. Category frequencies were transformed to percentages relative to the total number of exercises in the target thematic units of each course.

Table 1. Definitions of Different Levels of Cognitive Complexity

Category	Definition
Complex use of procedures	Task requires students to make meaning of a situation by applying a given set of concepts or statistical tools, in order to recognize patterns, reduce data, or estimate statistics of interest, or draw conclusions.
Simple use of procedures	Task requires students to apply a well-rehearsed algorithm, with no attention to the meaning of the outcome in the context of a problem situation.
Direct recall	Task requires students to reproduce previously learned facts, rules, or definitions.
No performance standards	Task requires students to freely consider a situation or explore interactions of statistical parameters, without setting performance standards.

3.3.3 Analysis of Interactivity

For the analysis of interactivity, we focused on the use of feedback to inform students' performance on knowledge assessment items and of learning objects (applets) within instructional explanations. We wanted to know not simply how many interactive learning objects the courses had, but also what pedagogical function they played and what kinds of interactions they afforded. In a preliminary analysis we identified three different kinds of objects: computational, representational, and exploratory. *Computational* objects yield statistics for a specified data batch. These objects simplify algorithmic procedures during problem solving, thus reducing cognitive load and allowing students to focus on conceptual issues. Examples are objects that yield critical values of t or chi-square distributions, normalize scores, or check the computation of summary statistics. The students' activity here amounts to entering data or parameters. *Representational* objects generate one or more representations (e.g., histograms, scatterplots) for a given dataset. These objects allow for representation coordination, as when a histogram and a box plot are displayed side by side and, by operating on one of the representations, students can check corresponding changes in the other representation. Interaction here consists mostly of noticing critical features and engaging in reading of displays. Finally, *exploratory* objects enable students to inspect statistical processes by changing parameters and checking their effects within a single representation. Examples are applets that illustrate the effects of between-group and within-group variability on the F statistic, and simulations of sampling distributions where students can modify sample size and number of samples. Interaction here consists of modifying parameters, noticing critical features, and making conceptual connections between components of processes.

To produce an estimate of trends or emphases in the pedagogical functions of interactive learning objects, we counted the number of computational, representational, and exploratory objects in the three selected thematic units. When a given object was used several times, it was counted once.

4. RESULTS AND ANALYSIS

4.1 Online Resources

Table 2 summarizes the online resources available in each course. As can be seen, there are rather typical resources such as applets, feedback, search engines to locate content, and glossaries. Courses differ in their provision of online note taking fa-

Table 2. Available Online Resources per Course

Resources	Courses: CyberStats	SurfStat	Psych3000	Intro Stat	SeeingStat	Investigating Statistics
Applets	√	√	√	√	√	√
Videos						√
Statistical software	√		√			√
Virtual labs			√			
Note-taking facilities	√			√		
Course map	√	√	√	√	√	√
Glossary	√	√		√	√	√
Search engine	√	√		√	√	√
Course management system	√		√			
Links to external sources	√	√				√
Electronic forums	√		√			
Multiple-choice questions	√	√	√	√		
Short answer questions	√		√		√	√
Feedback	√	√	√	√	√	√

cilities, integrated statistical software, course management systems for instructors, and links to external Internet sources. There are also differences in the use of electronic forums to support online discussions and group work, virtual labs, and type of knowledge assessment questions. CyberStats, Psych3000, and Investigating Statistics stand out for their completeness.

4.2 Authenticity and Variability of Examples

Figure 1 shows the mean number of examples and authenticity index per unit and course. CyberStats and Seeing Statistics rank the highest in both the number of examples and the authenticity of example scenarios. CyberStats includes a high number of thorough examples with realistic datasets and research descriptions (e.g., Forbes' estimation of barometric pressure through the boiling point of water, relationship between the chirp of crickets and air temperature, relationship between calories and meat composition of hot dogs). This is particularly evident in the units on one-way analysis of variance and regression. It is not the case, however, in the unit on measures of central tendency, which includes presumably authentic datasets on temperatures in different cities, yet in a way not explicitly connected with any research question or general query. Although explanations in CyberStats rely on a variety of realistic examples, which illustrate different aspects of the target topics, Seeing Statistics takes a different approach: The bulk of the explanation rests on a combination of a single, cursory example (which maximizes tractability of procedures), frequent illustrative applets, and additional, more authentic examples as asides (students can inspect them at particular locations by clicking on one of the navigational icons). Some of these examples, which are selected from psychology, economics, engineering, and biology, include references to primary sources, are fully described, and constitute a refreshing intermission in the dullness of traditional examples referring to student scores on statistics classes.

The remaining four courses have lower means of examples per unit and rank lower on the authenticity index, although there is variation from one thematic unit to another. Psych3000, a courseware characterized by very well crafted explanations, in-

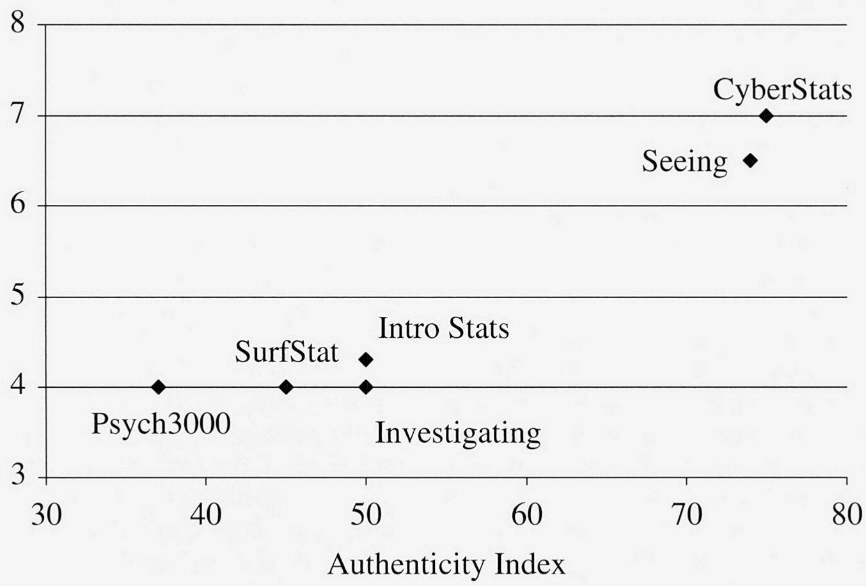

Figure 1. Mean number of examples and authenticity index per unit and course.

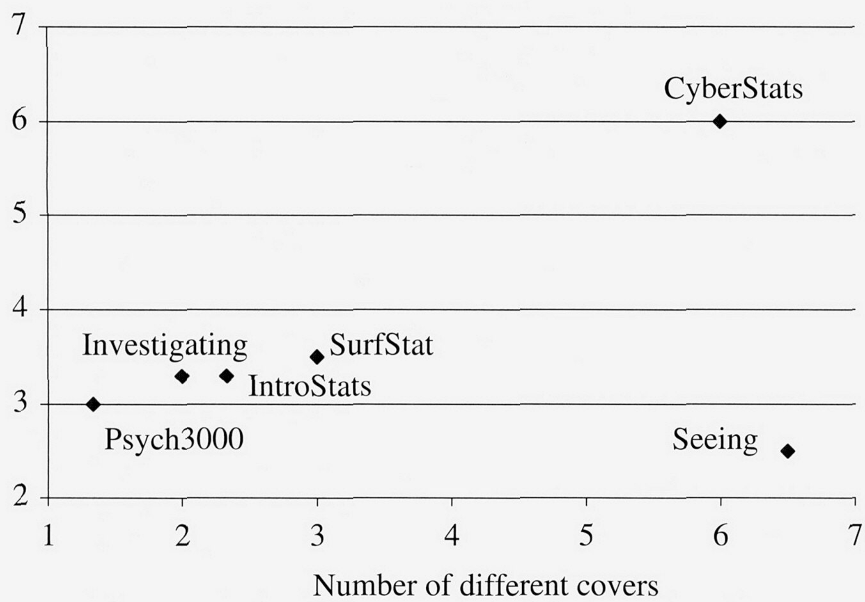

Figure 2. Mean number of different example covers and conceptually distinct examples per unit and course.

cludes very few examples (the majority of them with made-up datasets), which are compensated to some degree with fictitious descriptions that underscore critical issues of design, measurement, and inference. Again, the approach seems to maximize tractability at the expense of the authenticity of datasets.

Intro Stats, SurfStat, and Investigating Statistics fall approximately at the center of the scale of authenticity, with four examples per unit on average. Although some examples provide a realistic sense of datasets, more often than not example scenarios are poorly described. The statements can be as concise as the following two examples on regression from SurfStat:

1. The following table shows the rate of growth (%) of health expenditure per person in Australia at constant 1984–1985 prices (table and analyses follow).
2. To examine the effects of storage conditions on shelf life of a product several experiments were conducted. One involved measuring the moisture content of samples of the product at different levels of relative humidity (results and analyses follow).

The datasets and research settings referred to in these examples are plausible. Yet, their descriptions fail to locate the specific relation examined in a larger context of inquiry, and also fail to make a compelling case for the significance of the data. Additionally, issues of design and measurement are not mentioned. In this way, the practice of statistics is detached from the larger frame of scientific activity and decision-making.

Psych3000, SurfStat, Intro Stats, and Investigating Statistics are prolific in their use of "syntactic" examples, that is, datasets with no cover stories. For example, Intro Stats' unit on measures of central tendency includes only one content-based example (a dataset on shoe sizes, shoe widths, and gender) out of a total of nine examples. The remaining examples amount to a series of bare numbers that instantiate algorithms for the mode (in unimodal and bimodal distributions), the median (for even and odd datasets), and the mean. Although bare datasets with few data points make statistical patterns more salient, they strip statistical analysis from meaning making and inferential statistics from matters of practical significance. They also assume that what is "difficult" for students is the computational activity while what seems far harder is to think statistically with tools that are put to appropriate use (Lovett 2001; Lovett and Greenhouse 2000).

Figure 2 displays the mean number of cover stories against the mean number of different topics per course. Courses in the upper-right section of the display have sets of examples within thematic units that are comparatively variable both in their covers and topics. Courses in the lower-right area tend to have less variable sets of examples. On this dimension, CyberStats again ranks high. CyberStats provides examples in areas such as meteorology, public policy, biology, physics, and education. Its examples illustrate core concepts and procedures, but also critical conceptual issues (e.g., effect of outliers on measures of center, effects of the shape and spread of a data distribution on the F statistic, distinction between statistical and practical significance) which may not be salient enough when only prototypical cases are given.

Seeing Statistics offers a different kind of variability. There is an effort to provide students with examples in domains that they can relate to (psychology, biology, economics, and engineering). However, most of the examples use datasets with similar distributional properties, so examples are equivalent in terms of types of inferences they allow. The remaining courses rank comparatively low in both content and conceptual variability, and rely on fewer and more homogeneous examples.

4.3 Number, Authenticity, and Complexity of Exercises

Figure 3 displays the average number of exercises and the average authenticity index per unit in each course. Psych3000 and CyberStats have the greater average number of exercises per unit. Although it does not rank high on the exercise authenticity index, Psych3000 provides students with plenty of opportunities for practice, both through printable homework problems and applets that randomly generate datasets every time that the student runs the practice applets. CyberStats includes relatively

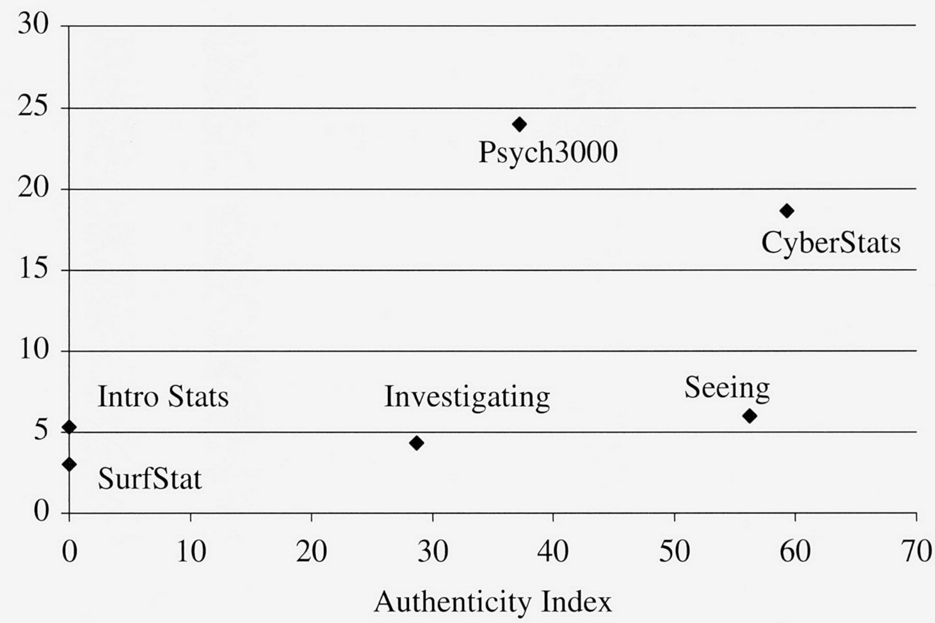

Figure 3. Mean number of exercises and authenticity index per unit and course.

authentic practice exercises, which are located next to the relevant instructional explanation (embedded problem solving) and at the end of each unit.

Including a small number of practice exercises may suggest that designers expect the courseware to be complemented with additional learning opportunities in the context of blended instruction. That may well be the case of Investigating Statistics, which is an online courseware explicitly developed for an on-campus course. However, it may also reflect the erroneous conception that understanding text explanations and engaging in online interactivity (in the context of exploratory applets) translate directly into skillful problem solving.

The distribution of levels of cognitive complexity across courses suggests that different courses make different assumptions as to what is meaningful in statistics learning. Some courses (Seeing Statistics) underscore conceptual understanding and, to that end, provide rich computational and representational support. Other courses (Psych3000) yoke conceptual understanding and knowledge of procedures through a thorough examination of mathematical definitions and derivations. These and other approaches to statistics learning express themselves in the quality of the exercises given to students. The attributes of exercises are also dependent on pedagogical constraints. Complex problem solving tasks are costly in that they are taxing both to design and to solve. Thus, we might expect courses to include a comparatively low number of these tasks (at least per thematic unit). In contrast, computational problems are relatively easy to generate and, if appropriate conditions are given, students can promptly get proficient at solving them.

Figure 4 shows the percentage of exercises per course assigned to each level of cognitive complexity. As can be seen, there are courses such as CyberStats and Seeing Statistics whose practice exercises include the entire variety of complexity levels, with about 60% of their tasks involving complex use of procedures. An example of complex use of procedures is the Exercise 5 (Unit E-3, Questions 39–47) from CyberStats:

For her class project, a student in the author's "Design and Analysis of Experiments" course compared the workload at a Customer Service counter on different days of the week. Since she worked at that department store, she was able to obtain data on randomly selected days. The following table gives the number of customers served per day at the Customer Service counter [Table is shown]. Use the next WebStat Interactive with "Workload Customer Service" data loaded. 39- Describe the central statistical question in the context of problem. 40- Identify the null and alternative hypotheses to be tested. 41- Describe different sources of variation in these data. 42- Why is this a one-way analysis problem? 43- Compare days using an appropriate comparative graph. What do you notice about the differences in number of customers served on different days? 44- Calculate descriptive statistics for each day. What do these statistics indicate about the differences among days of the week? 45- Use WebStat to compute the test statistic and the p value. 46- Use this p value to state a conclusion. Is there a significant difference in the mean number of customers served on different days of the week? 47- Do you think day of the week is a major contributor to the variation in the number of customers served? Why or why not?

In this exercise, students have the opportunity to reflect on issues of design, describe data, select relevant data displays, run statistical analyses, and engage in inference making and explanation; that is, they can participate in a meaningful array of statistical practices.

Courses that emphasize complex use of procedures contrast with courses, such as Intro Stats and SurfStat, which give no opportunities for practice and only include self-assessment items aimed at memorization of statistical facts and concept definitions (e.g., "the median of a distribution is: a) a measure of its dispersion; b) a measure of its location; c) its center of mass"). Somewhat in the middle is Psych3000, which provides numerous opportunities to practice computations (e.g., "Find the mean and mode of each of the following sets of measurements: a) 10, 8, 5, 0, 8, 3, 2, 5, 8, 2, 0; b) 1, 3, 3, 5, 5, 5, 7, 7, 9; c) 127, 7, 3, 0, 5, 3, 2, 3"). Yet these opportunities are complemented with relatively few, though meaningful, problem-solving activities in the context of a virtual lab. Virtual lab activities require students to engage in a significant variety of intellectual and decision-making practices, typical of statistics as an applied discipline, from examining research questions and design to selecting sta-

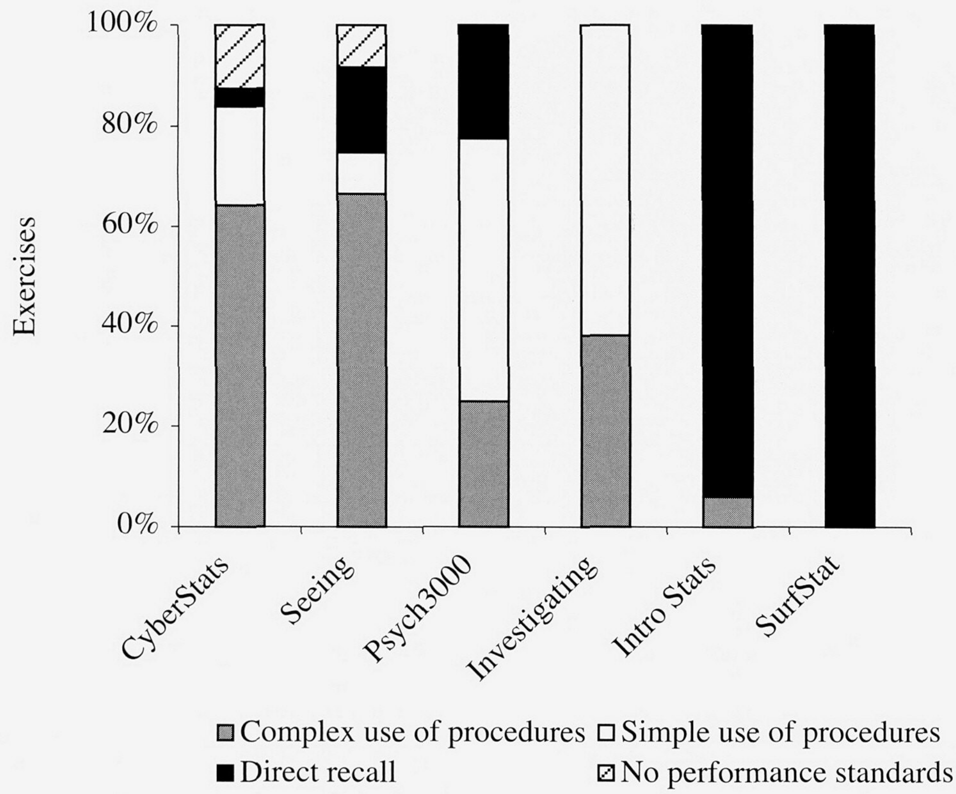

Figure 4. Percentage of exercises per course assigned to different levels of cognitive complexity.

tistical tools and analyses and drawing conclusions from evidence. Investigating Statistics adopts a similar approach, combining purely computational tasks with more ambitious course projects.

4.4 Analysis of Interactivity

Frequencies per course and kind of interactive object are reported in Table 3. Results show CyberStats with the highest number of interactive learning objects. Seeing Statistics follows, and then, respectively, Investigating Statistics, Psych3000, Intro Stats, and SurfStat. Yet, while the frequency of interactive objects is indicative of the overall reliance on this type of artifacts, the relationship between different types of objects and different styles of explanation is more informative. CyberStats is a course that strives for a balance between conceptual understanding and knowledge of procedures, in that concepts are brought as rationales for computations and procedures are taught to illustrate where numerical measures come from. Yet, CyberStats relieves students from hand computations. Interactive objects are comparably used to support problem solving (i.e., computational) and as argumentative backings or learning devices (i.e., exploratory and representational). In contrast, Seeing Statistics relies heavily on exploratory objects as visual illustrations of explanatory statements. Computational formulas and mathematical definitions are introduced as an aside without the support of interactive objects, which accounts for the higher percentage of exploratory applets relative to computational ones. This reliance on interactive objects for explanatory purposes also occurs in the unit of regression of Investigating Statistics, which rests heavily on two applets. In turn, instructional explanations in Psych3000 use computationally simple examples and slowly walk students through both mathematical and computational definitions. Most importantly, text explanations are self-contained in that they do not require interactive objects to be complete and fully understandable. Applets movies constitute an additional learning experience, not an indispensable component of the explanation. As sources of interactivity, Psych3000 resorts also to instructional games and virtual lab assignments. These features account for the comparatively low use of interactive objects in Psych3000. Applets in Intro Stats and SurfStat play a very marginal, almost dispensable, role.

4.5 Sampling Distributions: A Privileged Example of Interactivity and Visualization

The affordances of interactivity and visualization through applets are put to the test when we turn to highly theoretical con-

Table 3. Frequencies (and percentages) of Different Kinds of Applets per Course in Selected Units

Applet type	*CyberStats*	*Seeing Statistics*	*Psych3000*	*Investigating*	*Intro Stats*	*SurfStat*
Computational	11 (28)	1 (3)	1 (20)	6 (46)	1 (100)	0 (0)
Representational	10 (26)	3 (8)	2 (40)	1 (8)	0 (0)	0 (0)
Exploratory	18 (46)	33 (89)	2 (40)	6 (46)	0 (0)	2 (100)
Total	39 (100)	37 (100)	5 (100)	13 (100)	1 (100)	2 (100)

cepts and principles that are counterintuitive in the sense of not being easily related to everyday experiences. In such cases, exercises may be not relevant because learning may not entail skill acquisition, and examples may be hard to come by because of the abstract, theoretical nature of notions. One such concept is sampling distribution, which is foundational to inferential statistics and one of the most challenging topics for students to learn. Sampling distribution is a cognitively demanding concept because it refers to a theoretical distribution (result of an infinite number of samples), involves meaningful conceptual distinctions (e.g., distribution of scores versus distribution of statistics), and requires a solid grasp of sampling variability, a core concept that has shown to be a source of robust student misconceptions (Chance, delMas, and Garfield 2004).

How do the selected courses compare in terms of their treatment of sampling distribution and their use of applets to support instructional explanations? Overall, differences are more a matter of emphasis than a matter of substance. All courses link the notion of sampling distribution to the issue of accuracy of parameter estimates and sampling error; emphasize the distinction among sampling, sample and population distributions; and list core mathematical properties of the sampling distribution of the mean. Some courses (CyberStats, Investigating Statistics, Intro Stats, Psych3000) devote exclusive thematic units to sampling distributions, while others (Seeing Statistics, SurfStat) introduce the concept in the context of hypothesis testing for given statistics.

Differences also lie in the explanatory connections to data modeling and in the use of interactive learning objects. With respect to data modeling, Psych3000, unlike other courses, introduces the concept of sampling distribution in the context of the process of scientific inquiry. This approach highlights the role of sampling distributions in both statistical inference and scientific generalization and accuracy. In this way, the concept of sampling distribution is not reduced to a mathematical curiosity of pale significance compared to that of standard error, critical values, and confidence intervals. In terms of interactivity, all courses, except SurfStat and Intro Stats, use interactive learning objects to illustrate the process of iterative sampling from a population and to display dynamic sample and sampling frequency polygons. Some interactive objects (CyberStats and Seeing Statistics) focus on changes in the distribution of statistics as a function of sample size. Others (Psych3000) allow for representation coordination. Students can focus on critical features within a single representation (e.g., shape, skewness) and compare a single feature across multiple representations (sample, population, and sampling histograms). In terms of the usability, Investigating Statistics falls short because the simulation of sampling distribution actually requires using statistical software (Statlets) to sample from a randomly generated set of data, compute summary statistics, repeat this process n times, and finally plot the results.

Differences aside, these applets constitute typical examples of representational interactive objects, which generate multiple representations at a very low cost, freeing students from the practicalities of data sampling and display construction, and allowing them to attend to critical attributes of the resulting distributions. Yet it remains an empirical question the extent to which students can achieve deep understanding of sampling distributions only through learning by doing and discovery, with minimal direct instruction.

4.6 Virtual Labs

Interactivity is built in the selected courses mostly through interactive learning objects (Java applets and Flash movies) that illustrate specific concepts and procedures. Virtual labs, where students engage in extended problem solving activities and experience a wider range of statistical practices, are much less frequent. Psych3000 stands out as an exceptional case, both in terms of the variety of interactive environments it includes, and the care with which they have been designed and pedagogically documented. [Psych3000 is, incidentally, the only online course that provides access to scientific documentation regarding learning effectiveness and design principles (Malloy 2001; Malloy and Jensen 2001).] Psych3000 involves two "instructional games" ("Detect the difference" and "Difference to inference"), where students solve puzzles that require estimating the significance of mean differences. Students are expected to use these games as integrative activities after units on variability, regression, interactions, and correlation. Psych3000 also includes a virtual lab, where students conduct research on psychological issues within a simulated reality. Students read literature relevant to the research question, select research designs, define variables, select statistical analyses, and report results and interpretations. Although the importance of instructional environments such as these cannot be overemphasized, we see as limitations of the Psych3000 environments the lack of adaptive scaffolding and tutoring (through feedback), and the unnecessary and limiting artificiality of the simulated reality in which research questions are formulated.

4.7 Feedback

The use of feedback is highly uniform across the reviewed courses. If we exclude the responsive nature of some representational and exploratory applets, feedback is limited to supplying correct answers to open-ended questions (for the students to compare) and to the automatic grading of multiple-choice items. Feedback is given in the context of knowledge assessment activities, usually at the end of a thematic unit. The only course that provides feedback throughout is CyberStats, which customarily includes opportunities for practice immediately after sections of content explanations.

Although providing correct answers to questions and grading responses to multiple-choice items allow students to monitor their performance, two major limitations are worth mentioning. First, feedback is optional in that students need to click on an icon to get it. This optionality may hinder learning when students overestimate the accuracy of their knowledge, as usually happens with beginning learners (Rozenblit and Keil 2002). Second, feedback is not adaptive in the sense that it is not constructed vis-a-vis the particulars of the student's answer and mental model. So, while the students are given access to correct answers, feedback may be still too general to debunk specific misconceptions or debug faulty procedures. Unfortunately, providing adaptive feedback requires more sophisticated programming and authoring technologies, as well as detailed cognitive analyses of task environments and student learning processes

(Aleven and Koedinger 2002; Koedinger, Aleven, and Heffernan 2003; Murray 1999).

5. CONCLUSIONS

In this final section, we would like to make some general remarks on the courses and advance some suggestions for future directions in statistics online education. However, two caveats are in order. First, we understand that there is a timeboundedness to this examination. Online technologies change fast and newer versions of the reviewed courses are likely to incorporate features that further support interactivity and representation. These new online affordances may change what gets explained and how it is explained. Likewise, as statistics education becomes ubiquitous in elementary and secondary education, online courseware will surely adapt to changes in the students' entry competences, most likely going in the direction of greater conceptual density and depth. Second, although we have looked into critical features of instructional explanations and learning environments, we conceive a thorough evaluation of online courseware as relying upon empirical research on student learning, analyses of instructional explanations, considerations of the nature of statistical knowledge, usability analyses, and examination of the conditions of use.

In the topics we explored in greater detail (measures of central tendency, sampling distributions, simple linear regression, and one-way analysis of variance), we did not detect any glaring conceptual errors or misleading accounts. If anything, there is a visible effort in most courses to support conceptual understanding through visualization and explanations that unravel the meaning of measures and procedures. However, explanations vary in terms of their reliance on interactive learning objects, how explanations of concepts are coupled with mathematical derivations or computational procedures, and the attention devoted to the conditions of use of statistical measures and procedures.

Differences are also noticeable in the use of examples. We insisted on examples because they not only help students instantiate abstract concepts and illustrate procedures, but also convey a sense of the problems and issues that matter to the discipline. This is what we refer to as authenticity. Although some of the reviewed courses (CyberStats and Seeing Statistics) score high on authenticity, we believe that there is still considerable room for improvement. Perennial examples and exercises on student scores on tests in a statistics class, students' heights and weights, and decontextualized SAT or GRE scores need to be edited out and replaced with datasets derived from questions whose answers do make a difference. Likewise, example and problem statements need to address (however briefly) issues of design, measurement, and data structure, and how such issues are linked to disciplinary matters. Richer descriptions not only make a stronger case for the interestingness of datasets, but also make the mapping of statistical concepts and processes onto disciplinary contents more transparent and accessible to students. We also believe that online courses (and classroom instruction, as well) need to provide environments where students make decisions at different moments in the process of data modeling, from decisions about the resolution of measures and the plausibility of outcomes to decisions about the tenability of statistical assumptions and inferences. In that sense, virtual labs that support integrative problem solving are greatly recommended.

With respect to interactivity, some of the courses have come a long way from traditional printed materials. Not only is there extensive use of hyperlinks (which allow students to bridge knowledge gaps and follow side explanations), but also representational capabilities have been greatly enhanced. Yet, interactivity still needs to go beyond animated representations and provision of ready-made feedback. In the future, it needs to be complemented with adaptive feedback and scaffolding tailored to the students' individual needs. In this, we are far from meeting the expectations and potentials of online instruction. Online instruction is endowed with affordances that can potentially reconcile the goal of broad educational coverage and the ideal of individualized instruction. Success in such an endeavor depends on the design of materials that are easy to use, intellectually compelling, and adaptable to individual learning goals. The examined courses satisfy these criteria in varying degrees.

[Received August 2004. Revised February 2005.]

REFERENCES

Aleven, V., and Koedinger, K. R. (2002), "An Effective Metacognitive Strategy: Learning by Doing and Explaining with a Computer-Based Cognitive Tutor," *Cognitive Science*, 26, 147–179.

Anderson, J. A. (1993), *Rules of the Mind*, Hillsdale, NJ: Erlbaum.

Anderson, J. R., Farrell, R., and Sauers, R. (1984), "Learning to Program in LISP," *Cognitive Science*, 8, 87–129.

Anderson, J. R., Greeno, J. G., Kine, P. J., and Neves, D. M. (1981), "Acquisition of Problem Solving Skill," in *Cognitive Skills and Their Acquisition*, eds. J. R. Anderson, Hillsdale, NJ: Erlbaum, pp. 191–230.

Chance, B., delMas, R., and Garfield, J. (2004), "Reasoning About Sampling Distributions," in *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, eds. D. Ben-Zvi and J. Garfield, Dordrecht, The Netherlands: Kluwer.

Cobb, G. W. (1987), "Introductory Textbooks: A Framework for Evaluation," *Journal of the American Statistical Association*, 82, 321–339.

Dunbar, K. (1998), "Problem Solving," in *A Companion to Cognitive Science*, eds. W. Bechtel and G. Graham, Oxford: Blackwell, pp. 289–298.

Fergusson-Hessler, M. G., and deJong, T. (1990), "Studying Physics Texts: Differences in Study Processes Between Good and Poor Performers," *Cognition and Instruction*, 7, 41–54.

Fisher, R. A. (1936), "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, 7, 179–188.

Funkhouser, H. G. (1937), "Historical Development of Graphical Representation of Statistical Data," *Osiris*, 3, 269–404.

Gibson, J. J. (1977), "The Theory of Affordances," in *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, eds. R. Shaw and J. Bransford, Hillsdale, NJ: Erlbaum, pp. 67–82.

Greeno, J. G. (1994), "Gibson's Affordances," *Psychological Review*, 101, 336–342.

Koedinger, K., Aleven, V., and Heffernan, N. (2003), "Toward a Rapid Development Environment for Cognitive Tutors," in *Artificial Intelligence in Education: Shaping the Future of Learning Through Intelligent Technologies, Proceedings of AI-ED*, ed. J. Kay, Amsterdam: IOS Press, pp. 455–457.

Kosslyn, S. (1989), "Understanding Charts and Graphs," *Applied Cognitive Psychology*, 3, 185–225.

Kulik, J. A., and Kulik, C. I. C. (1988), "Timing of Feedback and Verbal Learning," *Review of Educational Research*, 58, 79–97.

Langley, P., and Simon, H. A. (1981), "The Central Role of Learning in Cognition," in *Cognitive Skills and Their Acquisition*, ed. J. R. Anderson, Hillsdale, NJ: Erlbaum, pp. 25–45.

Larkin, J. H., and Simon, H. A. (1987), "Why a Diagram is (Sometimes) Worth Ten Thousand Words," *Cognitive Science*, 11, 65–99.

Lavinger, N. C., and Lajoie, S. P. (1995), "Learning Statistics Through Exemplars," in *Proceedings of the 17th Annual Conference of the Cognitive Science*

Society, eds. J. D. Moore and J. F. Lehman, Hillsdale, NJ: Erlbaum, pp. 643–647.

LeFevre, J., and Dixon, P. (1986), "Do Written Instructions Need Examples?" *Cognition and Instruction*, 3, 1–30.

Leinhardt, G. (2001), "Instructional Explanations: A Commonplace for Teaching and Location for Contrast," in *Handbook of Research on Teaching* (4th ed.), ed. V. Richardson, Washington, DC: American Educational Research Association, pp. 333–357.

Leinhardt, G., Zaslavsky, O., and Stein, M. K. (1990), "Functions, Graphs, and Graphing: Tasks, Learning, and Teaching," *Review of Educational Research*, 60, 1–64.

Lovett, M. (2001), "A Collaborative Convergence on Studying Reasoning Processes: A Case Study in Statistics," in *Cognition and Instruction: 25 Years of Progress*, eds. S. Carver and D. Klahr, Mahwah, NJ: Erlbaum, pp. 347–384.

Lovett, M., and Greenhouse, J. (2000), "Applying Cognitive Theory to Statistics Instruction," *The American Statistician*, 54, 196–206.

Malloy, T. E. (2001), "Difference to Inference: Teaching Logical and Statistical Reasoning Through Online Interactivity," *Behavior Research Methods, Instruments, and Computers*, 33, 270–273.

Malloy, T. E., and Jensen, G. C. (2001), "Utah Virtual Lab: JAVA Interactivity for Teaching Science and Statistics Online," *Behavior Research Methods, Instruments, and Computers*, 33, 282–286.

McCarthy, K. (2002), "Learning Genre From Examples: The Effects of Number of Examples and Scaffolding on Student Understanding and Writing," unpublished PhD dissertation, University of Pittsburgh.

McKendree, J. (1990), "Effective Feedback Content for Tutoring Complex Skills," *Human Computer Interaction*, 5, 381–413.

Mitchell, T., Keller, R., and Kedar-Cabelli, S. (1986), "Explanation-Based Generalization: A Unifying View," *Machine Learning*, 1, 47–80.

Murray, T. (1999), "Authoring Intelligent Tutoring Systems: An Analysis of the State of the Art," *International Journal of Artificial Intelligence in Education*, 10, 98–129.

NCTM (2000), "Principles and Standards for School Mathematics," Reston, VA: National Council of Teachers of Mathematics.

Pirolli, P. L. (1991), "Effects of Examples and Their Explanations in a Lesson on Recursion: A Production System Analysis," *Cognition and Instruction*, 8, 207–259.

Quillici, J. L., and Mayer, R. E. (1996), "Role of Examples in How Students Learn to Categorize Word Problems," *Journal of Educational Psychology*, 88, 144–161.

Rissland, E. L. (1989), "Example-Based Reasoning," in *Informal Reasoning in Education*, eds. J. F. Voss, D. N. Perkins, and J. W. Segal, Hillsdale, NJ: Erlbaum, pp. 187–208.

Rozenblit, L., and Keil, F. (2002), "The Misunderstood Limits of Folk Science: An Illusion of Explanatory Depth," *Cognitive Science*, 26, 521–562.

Shaffer, D. W., and Resnick, M. (1999), " 'Thick' Authenticity: New Media and Authentic Learning," *Journal of Interactive Learning Research*, 11, 195–215.

Shah, P., and Hoeffner, J. (2002), "Review of Graph Comprehension Research: Implications for Instruction," *Educational Psychology Review*, 14, 47–69.

Simon, H. A., and Zhu, X. (1988), "Learning Mathematics from Examples and by Doing," *Cognition and Instruction*, 4, 137–166.

Stein, M. K., and Lane, S. (1996), "Instructional Tasks and the Development of Student Capacity to Think and Reason: An Analysis of the Relationship Between Teaching and Learning in a Reform Mathematics Project," *Educational Research and Evaluation*, 2, 50–80.

Sweller, J., and Chandler, P. (1994), "Why Some Material is Difficult to Learn," *Cognition and Instruction*, 12, 185–233.

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.

Utts, J. (2003), "What Educated Citizens Should Know About Statistics and Probability," *The American Statistician*, 57, 74–79.

Wender, K. F., and Muehlboek, J. S. (2003), "Animated Diagrams in Teaching Statistics," *Behavior Research Methods, Instruments, and Computers*, 35, 255–258.