

Constructing a Computer-Assisted Testing and Evaluation System on the World Wide Web—The CATES Experience

Chien Chou, *Associate Member, IEEE*

Abstract—This paper describes applications of computer network technologies to testing and evaluation; reviews related research on computer-assisted testing; and introduces the analysis, design, progress-to-date, and evaluation of the Computer-Assisted Testing and Evaluation System (CATES) currently under development at National Chiao Tung University, Hsinchu, Taiwan, R.O.C. CATES shows how networks and the World Wide Web (WWW) are being used in science and engineering education. Technical implications of the study are also discussed.

Index Terms—Computerized testing, network-based testing and evaluation, World Wide Web application.

I. INTRODUCTION

THE emergence and continued growth of computer network technologies are changing the way people around the world work and learn. Computer networks offer new alternatives for creating, storing, accessing, distributing, and sharing learning materials. Moreover, computer networks provide new channels for interactions between teachers and students, teachers and teachers, and students and students. Therefore, what and how to apply these new technologies to engineering and science education is a great challenge for teachers, researchers, and instructional design professionals.

Among the many Internet applications, the World Wide Web (WWW) is one of the most popular because of its graphical user interface (GUI) design and equal access to available information resources. Many studies [1]–[3] have reported research results on the design and development of Web-based instruction. These reports show the exciting promise Web-based instruction provides, and its expected immense impacts on conventional classrooms as well as on the distance learning environment. As Alexander [4] stated, the Web provides an opportunity to develop previously unattainable new learning experiences for students.

Assessment is an essential aspect of all instruction. Teachers need to know what and how well students have learned, and so do students themselves. Assessment can take the form of a quiz or examination to test students' learning achievements, or of a questionnaire to investigate students' attitudes and

reactions to new instructional courseware. Assessment using Web-based networks is definitely an essential part of network learning. However, with students dispersed at distant sites and having different Web access opportunities, assessing student performance and attitudes on the Web may pose technical and logistical problems regarding the construction, distribution, collection, and administration of tests.

The main focus of this paper is how computer networks in general, and the Web in particular, are being used for testing and evaluation in Taiwan. This paper describes the Computer-Assisted Testing and Evaluation System (CATES) as an innovative application of network technologies to engineering and science education. The paper first reviews the literature on computer-assisted testing, then introduces the analysis, design, development, and evaluation of the CATES. Conclusions and plans for future work are also provided.

II. OVERVIEW OF COMPUTER-ASSISTED TESTING AND EVALUATION

A. Computer-Based Testing Versus Paper-and-Pencil Testing

The use of computers for testing purposes has a history spanning more than 20 years. The advantages of administering tests by computer are well known and documented, and include reduced testing time, increased test security, and provision of instant scoring (see [5]–[7]).

In early studies, the main research focus was on whether computer-based tests were equivalent to paper-and-pencil tests when computers gave exactly the same tests as those given in paper-and-pencil formats. In order to define score equivalence, the American Psychological Association (APA) in 1986 published the Guidelines for Computer-Based Tests and Interpretations. The guidelines define the score equivalence of computerized tests and conventional paper-and-pencil tests as 1) the rank orders of scores of individuals tested in alternative modes closely approximating each other and 2) the means, dispersions, and shapes of the score distributions being approximately the same, or capable of being made approximately the same by rescaling the scores from the computer tests versions [8]. The guidelines also require that any effects due to computer administration be either eliminated or accounted for in interpreting scores. In their empirical study, Olsen *et al.* [9] compared paper-administered, computer-administered, and computer-adaptive tests by giving third- and sixth-grade students mathematics applications achievement tests. This study

Manuscript received February 1, 1998; revised December 21, 1999. This work was supported by the National Science Council, Taiwan, R.O.C., under Project NSC87-2511-S-009-006-ICL.

The author is with the Institute of Communication Studies, National Chiao Tung University, Hsinchu 30010, Taiwan, R.O.C. (e-mail: cchou@cc.nctu.edu.tw).

Publisher Item Identifier S 0018-9359(00)03987-X.

found no significant differences between paper-administered and computer-administered tests, and equivalences among the three test administrations in terms of score rank order, means, dispersions, and distribution shapes.

Mazzeo and Harvey [10] pointed out that computer-based test graphics may affect test scores and consequently their equivalence with paper-and-pencil versions, and that tests with reading passages may be more difficult when given on computers. Bunderston *et al.* [5] suggested performance on some item types such as paragraph comprehension are likely to be slower if presented by computer, while some types such as coding speed items are likely to be faster.

In reviewing all the above-mentioned studies, Bugbee [11] concluded that the use of computers indeed affects testing; however, computer-based and paper-and-pencil tests can be equivalent provided the test developers take responsibility for showing that they are. Bugbee stated that the barriers to the use of computer-based testing are inadequate test preparation and failure to grasp the unique requirements for implementing and maintaining computer tests. In other words, Bugbee reminded us that some factors such as the design, development, administration and user characteristics must be taken into consideration when computers are used for testing.

B. The Use of Computers in Developing and Administering Tests

As computer-assisted instruction (CAI) has grown in popularity, computer-based testing has become more and more appropriate for assessing students' CAI learning achievement. As Bugbee [11] states, if what is being tested is done on or learned from a computer, then it is more appropriate to assess it by computer. Thus, computers are used as the sole vehicles for distributing tests, not only as alternatives to paper-and-pencil testing. Alessi and Trollip [12], in their classic book on computer-based instruction, devoted a chapter to the design, development, and use of computer-based testing. They pointed out that the two main ways of incorporating computers into the testing process are for constructing or administering tests. When constructing tests, test developers use computers' word processing abilities to write test items and use their storage capacities to bank and later retrieve test items. Jacob and Chase [13] pointed out that computers can present test materials paper-and-pencil test cannot, for example, 3-D diagrams in computer graphics, motion effects, rotating geometric forms, animated trajectories of rapidly-moving objects, and plants seen from different angles. Shavelson *et al.* [14] further suggest using computer simulations for hands-on performance assessment. In their project "Electric Mysteries," students were required to replicate electric circuits by manipulating icons of batteries, bulbs, and wires presented on a Macintosh computer.

When administering tests, computers can be used to provide individualized testing environments, that is, allowing students to take tests when they are ready. Moreover, test contents can be customized for students by providing different difficulty levels and emphases [12]. Computer-based testing can also be designed to provide test-takers with immediate feedback and

scoring. However, Wise and Plake [6] found that immediate feedback may contribute to students' test anxiety. Bernt *et al.* [15] also pointed out that general computer-test anxiety may influence test-takers. They considered that, although anxiety tends to be a random variable among people, it must be identified and dealt with. Jacob and Chase [13] also suggested discontinuing item-by-item feedback until further research has been done on the computer-test-anxiety issue.

C. The Use of Computer Networks in Testing and Evaluation

Advancements in computer networking technology have allowed stand-alone computers to be equipped with powerful communication abilities, thus providing an alternative for assessing students' learning achievements and attitudes. Students dispersed at distant sites may have options to take the test at different test locations and times. In addition to the traditional multiple-choice, fill in the blank, and short essay type questions, Rasmussen *et al.* [16] suggested Web-based instruction include participation in group discussions and portfolio development to evaluate students' progress. Khan [17] also suggested Web-based instruction designers have facilities that allow students to submit comments about courseware design and delivery.

Although many researchers, e.g., [16], [18], considered testing and evaluation to be of utmost importance in Web-based instruction and suggested some design strategies and techniques, few usable systems have been developed and no empirical data collected to explore the feasibility of computer-assisted testing and evaluation on the Web. The search for creative and effective tools and methods for conducting testing and evaluation in such a complicated technology-dependent learning environment represents a challenge for system designers and instructional designers. The innovative ideas presented in the CATES study are some first steps toward addressing this challenge.

III. THE CATES CASE

CATES is being designed to test students' learning achievements and evaluate courseware in Web-based learning environments. It is currently under development at National Chiao Tung University (NCTU), one of Taiwan's leading educational institutions devoted to instruction in science and technology. The CATES project is a team-oriented endeavor, involving five faculty members and over ten graduate students. In concept and construction, the CATES project is intended to integrate three major components: system, test items, and interface in its analysis, design, development, and evaluation phases.

A. Analysis Phase

When analyzing the use of a Web-based test, the dimensions of time and location of testing can help developers conceptualize system use. Noting whether they are specific or fixed, four sets of prepared tests, testing types, and situations can be characterized, as shown in Table I. The first set requires students to take a test at a fixed location (e.g., the computer center at NCTU) at a specific time, so it guarantees the

TABLE I
TEST TIMES AND LOCATIONS

	Test-taking time	Test location	Test sheet prepared	Type of test	Appropriate test situation
1	Specific	Fixed	One test	Open/closed book	College entrance examination, course midterm and final exam
2	Flexible	Fixed	Individual	Open/closed book	Driver license exam, school-wide placement test
3	Specific	Non-fixed	One test	Open/closed book	Midterm and final in distance education
4	Flexible	Non-fixed	Individual	Open book	Self assessment, practice

highest degree of testing fairness from the students' points of view. Appropriate testing situations for this set would be the college entrance examination and a course final examination. The second set requires students to take a test at a fixed location but the testing time is flexible. An appropriate testing situation would be the nationwide driver's-license examination or school-wide placement tests (such as in [19]). Test-takers can go to specific test locations when they feel they are ready. The third set requires students to take a test at a specific time but the testing location is not fixed. Students can even use their own computers to take the test. An appropriate test situation here would be a distance-learning midterm or final examination. The fourth set does not specify a test time or location; therefore, students can use the test for self-assessment and practice.

It is worth noting that in order to eliminate the possibility of students passing test information to others because of different testing times, such as in sets 2 and 4, different test sheets for individual students must be prepared. This can be done by composing sheets of items selected at random from a test bank. If test locations are not fixed, as in sets 3 and 4, open-book-type tests should be employed because supervision may not be available. From the users' viewpoint, testing flexibility in terms of time and location increases from set 1 to set 4. From the test developers' viewpoint, however, the difficulty of designing the test systems increases from set 1 to set 4. CATES is aimed at meeting the four situations.

B. Design Phase

CATES is built on the Internet. By using HTML, courseware authors and test developers can create hypertext nodes incorporating multimedia content and forward them to other nodes through links. In order to put the HTML test items on Web servers and collect responses, Web test developers must develop common gateway interface (CGI) programming, which is a standard that enables external gateway programs to interface with other information servers. CGI programming is usually written in Visual Basic (VB), C, or Perl. Web test-takers can employ browsing software such as Netscape or Internet Explorer (IE) to access and take tests.

The system design for CATES can be viewed as a three-layer structure, as shown in Fig. 1. The User Interface Layer consists of four modules: students take test, teachers construct test items, developers select existing test items from an item

bank, and teachers/evaluators access students' test results. The System Management Layer contains three management programs: Item bank manager, Test results manager, and Learning path and pattern recorder, and communicates with the first layer through activities such as presentation and logging. The Student Modeling Layer has two purposes: it constructs student models by analyzing learning paths and pattern results, then forms instructional analysis and guidance to adjust learning and testing environments through controlling the managers in the second layer.

Because this paper concentrates on presenting the interactive testing environment from the users'—students and teachers—points of view, the next discussions focus on the User Interface Layer. The client/server architecture was adopted for the system design. For example, the managers run on server machines and provide instructional and communication information and services. The interface programs, on the other hand, are done by client machines to fulfill the personal requests of individual students.

C. Development Phase

The development of CATES involved programming the specifications to fulfill the production requirements identified in the analysis and design phase. These include a working prototype system built on the Windows NT version 3.5 Chinese operating system using Web site version 1.1 as the web server software, and Visual Basic 4.0 for CGI programming. The test items, student information, and test results were stored in a database developed using Microsoft Access. Related technical developments are described in the section of Technical Implications.

The test-item bank contains four sets of items at present: educational technology, computer networks, multimedia systems, and basic computer concepts (BCC), consisting of 32, 25, 20, and 96 items, respectively. Included are multiple-choice, matching, short-answer, and essay questions. Because of the small number of test items currently available in the educational technology and computer networks tests, all students received exactly the same test sheets. In the case of the multimedia systems test, CATES randomly selected ten questions out of twenty which students had received as preparation directions before taking the network test, so the test questions for each student in this case was individualized. In the case of the BCC test, a reflective-type question was presented at the end of instruction on each Web page.

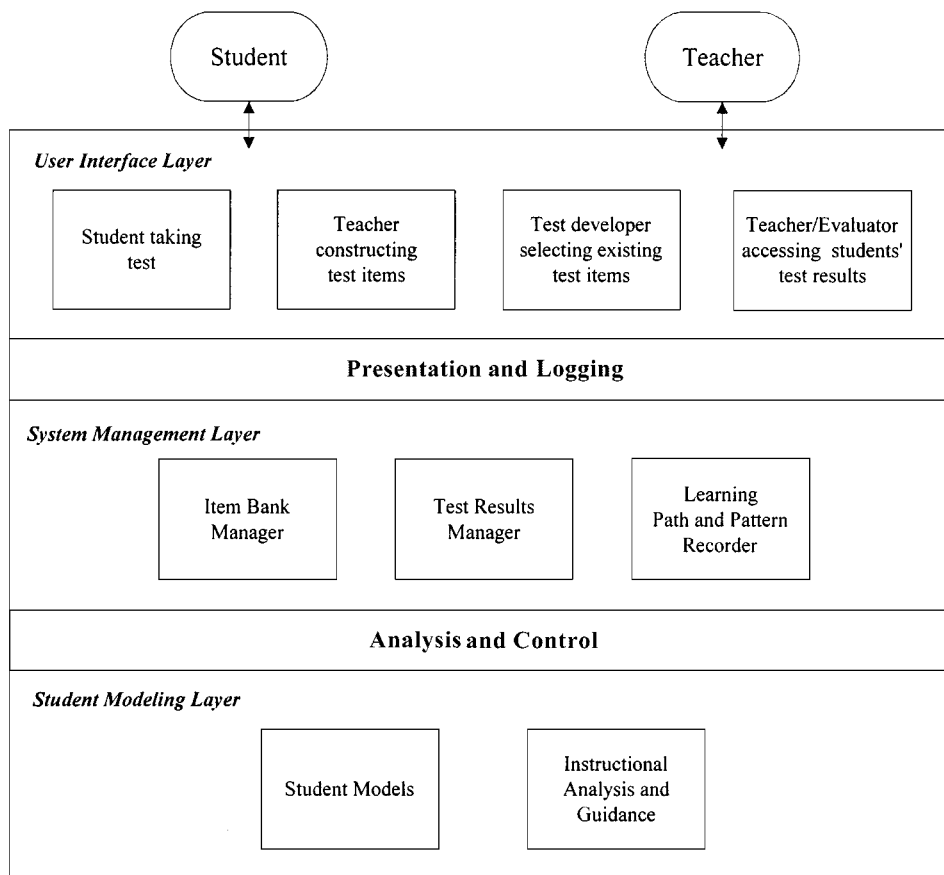


Fig. 1. CATES system structure.

An on-line questionnaire has also been developed to assess students' attitudes toward their Web tests. Students are asked whether they like, trust, or are anxious about the Web-based test. The questionnaire also asks them to compare Web-based testing with paper-and-pencil testing in terms of ease of reading, speed of responding, ease of cheating, and so on. The questionnaire is presented after testing but before students can check their multiple-choice and matching-question test results.

The test and questionnaire are presented in a "form" provided by the Web. Forms are tools used to gather information from Web users. In this current form, multiple-choice questions may have "radio buttons" that allow students to make the choices by pointing and clicking. Short-answer questions may present "text entry boxes" that permit full sentences and paragraphs to be typed in. Once students complete their tests and questionnaires, they then click on "submit" buttons to electronically send the entire form to the database at the server site.

Wise and Plake (1989) noted that computer-based testing usually does not include three features of paper-and-pencil testing: 1) allowing test takers to skip questions and come back to answer them later; 2) allowing test takers to review previously completed questions; and 3) allowing test takers to change previously entered answers. In this current test and questionnaire design, all items are placed on one web page. Although the page is more than one screen, students can scroll up or down to see different items. This allows them to skip questions, review pre-

viously completed questions, and change entered answers prior to submitting their tests and questionnaires.

D. Evaluation Phase

In addition to the system developers' self-testing (α -test), a formative evaluation was conducted to examine the usability of the CATES functions and the presentation of the test items. Two major evaluation approaches were adopted: expert-based and user-based [20]. One experienced computer programmer was invited to check the system's programming and functions. One interface expert was invited to evaluate how well the interface meets the goals of reducing fear of using the system and fear of making errors, and increasing the system's efficiency, usability, and user-satisfaction. One experienced test developer checked the way test items were presented on the computer screen. The system, interface, and test-item presentation were then debugged and revised according to the evaluation results.

The purpose of the user-based evaluation was also to explore the usability of the system, the presentation of the test, and the students' experience using computers, but not to examine students' learning achievement. Therefore, the difficulty, discriminability, reliability, and validity of the test itself are not presented as research foci in this paper. The user-based evaluation was conducted four times (Table II) by four classes:

- 1) 27 undergraduates participated in the first tryout, the educational technology test;

TABLE II
USER EVALUATION OF CATES

	Test-taking time	Test location	Number of students participating	Type of test	Topic of the test
1	Specific	Fixed (main computer center)	27 undergraduates	Closed book	Educational Technology
2	Flexible (9:00AM-5:00PM)	Fixed (main computer center)	15 graduates	Partial open book (random questions)	Multimedia Systems
3	Specific	Partially-fixed (main and branch computer centers with on-site supervision)	17 graduates	Closed book	Computer Networks
4	Flexible (while reading the instructional materials)	Non-fixed (at students' computers which delivered the instructional pages)	42 undergraduates	Open book (self practice—questions presented at the end of the reading materials)	Basic Computer Concepts (BCC)

- 2) 15 graduate students in the second tryout, the multimedia systems test;
- 3) 17 graduate students in the third tryout, the computer networks test;
- 4) 42 undergraduates in the fourth tryout, the basic computer concept self-practice.

The first two evaluations were conducted at the NCTU main computer center at a specific time. The third evaluation was conducted at the NCTU main, as well as at a branch computer center at a specific time. In the case of the fourth evaluation, the questions were presented at the end of each instructional Web page; the students could use their own computers to read and reflect on the questions while they proceeded on the learning journey. Since this was not a formal test, the author could only conduct post-hoc interviews to investigate students' opinions on the design of the questions, but not collect their answers to the questions. One of the reasons that these four classes were selected to participate in the CATES evaluation was that part of their classes had been based on computer networks, e.g., learning materials was distributed on the WWW, students' homework was submitted via E-mail, and BBS discussion groups had been formed. Therefore, the author felt it was appropriate to assess students' opinions about computer networks since what they had learned or been tested on had been done by computer.

Students reactions to the Web-based testing revealed positive but inconsistent results. From the first three groups of testers, a total of 59 students indicated similar attitudes toward their Web-based testing experience. About 70% of the students indicated they liked this kind of test, and about 67% considered Web-testing to be more efficient than paper-and-pencil testing. However, while 75% students thought the test was easy to read on the screen, about 67% said the Web test was more likely to tire them than paper-and-pencil testing. When asked whether Web-testing increased their test anxiety, about 30% answered yes, half answered no, 20% expressed no opinion, and none said it decreased their anxiety. If asked to choose between a Web test and a paper-and-pencil test, about 50% of the students said they would choose the Web test, 15% would choose paper-and-pencil, and about 35% indicated no preference. Table III lists the questionnaire items asked and 59 students' reactions.

TABLE III
QUESTIONS AND STUDENTS' REACTIONS

Question	Answer	Number of students (total = 59)	Percentage
I like the Web-based test.	Yes	41	69.49
	No	10	16.95
	No opinion	8	13.56
Web-testing is more efficient than paper-and-pencil testing.	Yes	40	67.79
	No	5	8.47
	No opinion	14	23.73
The Web test is easy to read on the screen.	Yes	44	74.58
	No	8	16.95
	No opinion	7	11.86
The Web test was more likely to tire me.	Yes	40	67.79
	No	11	18.64
	No opinion	8	16.95
The Web test increased my test anxiety.	Increased	18	30.51
	Not increased	29	49.15
	No opinion	12	20.33
	Decreased	0	0.00
The typing sound made by other people made me nervous.	Yes	20	33.90
	No	32	54.24
	No opinion	7	11.86
My typing speed significantly influenced my completion time.	Yes	30	50.85
	No	25	42.37
	No opinion	4	6.78
The choice between a Web test or paper-and-pencil one for my next test.	The Web test	30	50.85
	The paper-and-pencil test	9	15.25
	No preference	20	33.90

Note that about one-third of the students said that the sound of typing created by other testers made them a little nervous. About half of the students said the typing speed significantly influenced their completion times. It seemed that as more typing was needed, such as in the multimedia systems test which consisted of only ten randomly-selected essay questions, the typing speed and typing sound had a more serious impact on the students' responses. From on-site observations and follow-up interviews, two reasons were identified to explain this reaction. First, typing Chinese is, by nature, more difficult than typing English. Second, there are at least five ways users can type Chinese into computers, and one particular method requires a change in keyboard key labels. In both tryouts, students choosing to use the less popular Chinese typing methods were handicapped by the existing key labels provided by the computer centers. Although these students finally finished the test, their completion times

(average 135 min for the educational technology test, 150 min for the multimedia systems test, and 130 min for the computer networks test) were both far slower than those of other students (average 90 min for the educational technology test, 180 min for the multimedia systems test, and 100 min for the Computer Networks test), and slower than they themselves had expected.

Six graduate and five undergraduate students were randomly selected for interviews after their tests. The first question asked them to comment on their Web-testing experience. Most of the graduate students expressed that, since the network is part of their school lives, network-based testing in general and the Web test in particular would naturally become a part of their lives too. They all believe this kind of computerized-testing will be more and more popular in the future.

Three of the five undergraduates pointed out that they were more nervous about the test because of the test itself and the content (educational technology) than because of the medium (network). One of them said, "As a student, you never like a test and are nervous about it, no matter what."

When asked to think about the advantages of a Web-test from a test-taker's viewpoint, one graduate said the Web-test is environmentally-friendly: saving paper, ink, and poisonous liquid-paper. She used to collect all the tests she had taken since childhood, which occupied a lot of space. One student said he typed faster than he could write, and he believed his poor hand-writing often caused him to lose points. Three students agreed with his concern and believed that the hand-writing on a test influenced the scores to some degree. "The fact that one needs to type on computerized tests," they concluded after a short discussion, "would lead to fairer grading."

Students were asked about the weaknesses they saw in Web-tests. None of the students thought they were handicapped by the network or Web since the students are very computer-literate at this science- and technology-oriented university. Four students were concerned that, since Web-tests depend totally on computer networks, they are prone to technological problems. Although no problems occurred during their tests, they feared possible disconnections or breakdowns during their tests sometimes. This created another kind of test anxiety, namely, anxiety about the test medium. Therefore, a safer, more reliable, and more anxiety-free Web-test system and related research are warranted.

IV. TECHNICAL IMPLICATIONS

Although there are some commercial Web-based testing products available, most of them are in English, such as Quiz Wiz [21]. While each English letter requires a single byte in terms of computer memory, each Chinese character occupies two bytes. This makes it impossible for any English system to present Chinese test items or to take Chinese answers. In consideration of this special requirement for a testing system, a system in which both English and Chinese can be used simultaneously was developed. The CATES not only provides major functions such as test-taker's identification, response storage, automated scoring, test-version control and so on, as many

commercial products do, but also serves as an environment for our continuous research in educational and technical aspects.

The CATES has three major technical developments. The first one is that it is now built as a multi-server system—WebCAT*. WebCAT* provides facilities for testbase sharing, real-time test generation, and cross-testbase analysis. Through their own Web browser, each teacher can access other item bank questions stored in different servers and thus have a larger number of usable test items. Real-time test generation could produce different tests upon request, thus preventing cheating and making network testing more feasible. In order to providing insights into the performance of individuals and groups, cross-testbase analysis can accurately assess the difficulty, discriminability, reliability, and validity level for each test item and for each testbase as a whole. For detailed description of WebCAT*, please refer to [22].

The second development conducted by the CATES team is the development of the Computer Logging of User Entries (CLUE) sub-system. The CLUE is an innovative approach to formative evaluation of the courseware content in Web-based distance learning environment. CLUE combines the existing computer logging and widely used self-reporting methods to let courseware reviewers mark parts of contents they think should be revised. It provides pop-up windows for reviewers to enter their comments, and performs calculations of all input entries for course developers. Thus, CLUE saves much tedious recording work during the evaluation process. For detailed description of CLUE, please refer to [23].

The third technical advancement in CATES is the development of "adaptive questionnaires" to assess Web users' attitudes. Unlike past adaptive tests based on item-response theory (IRT) such as those used in computerized GRE and GMAT, the new adaptive questionnaires will use CGI programming to control presentation. Adaptive questioning uses the answers to certain questions to determine the next series of questions and to skip unrelated questions. For example, students answering "no" to question 5 are taken to question 11, while students answering "yes" are routed to complete questions 6–10. The process is used frequently in most surveys. However, Web questionnaires can be programmed so questions 6–10 will not be shown at all to the students answering "no" to question 5. Pitkow and Recker [24] showed that Web-based questionnaires can reduce the number and complexity of questions presented to users.

V. CONCLUSION

The CATES system the author has described is a collective and collaborative project intended to integrate an interactive testing system and research effort in concept and construction. This study provided a four-mode framework, based on the dimensions of time and location of testing, to help Web-based test developers conceptualize system use. This study also collected empirical data on these four different modes, and discussed the technical implications. As the world becomes more computer-network oriented and connected, so too must testing. While paper-and-pencil testing and students' test anxiety will never go

away, computerized testing in general, and Web-based testing in particular, are growing. From personal development and utilization experiences with CATES, the author concludes that Web-based computerized testing is becoming a major testing alternative. It is expected that more research will be conducted on the interactive Web-based testing environments and systems like CATES will be developed to benefit the students.

ACKNOWLEDGMENT

The author thanks Research Assistant H. Lin and Programmer C.-C. Tao for their contributions to this project.

REFERENCES

- [1] C. Chou and C. T. Sun, "Constructing a cooperative distance learning system: The CORAL experience," *Educ. Technol. Res. Develop.*, vol. 44, no. 4, pp. 85–104, 1996.
- [2] T. Reeve and P. M. Reeves, "Effective dimensions of interactive learning on the World Wide Web," in *Web-Based Instruction*, B. H. Khan, Ed. Englewood Cliffs, NJ: Educational Technology, 1997, pp. 59–66.
- [3] M. El-Tigi and R. M. Branch, "Designing for interaction, learner control, and feedback during Web-based learning," *Educ. Technol.*, vol. 37, no. 3, pp. 23–29, 1997.
- [4] Teaching and Learning on the World Wide Web, S. Alexander. (1995). [Online]. Available: <http://www.scu.edu.au/ausweb95/papers/education2/alexandr/>
- [5] C. V. Bunderson, D. K. Inouye, and J. B. Olsen, "The four generations of computerized educational measurement," in *Educational Measurement*, 3rd ed, R. L. Linn, Ed. New York, NY: Amer. Council Educ.—Macmillan, 1989, pp. 367–407.
- [6] S. L. Wise and B. S. Plake, "Research on the effects of administering tests via computers," *Educ. Meas.: Issues Practice*, vol. 8, no. 3, pp. 5–10, 1989.
- [7] A. C. Bugbee, *Examination on Demand: Findings in Ten Years of Testing by Computer 1982–1991*. Edina, MN: TRO Learning, 1992.
- [8] *Guidelines for Computer-Based Tests and Interpretations*. Washington, DC: Amer. Psychol. Assoc., 1986.
- [9] J. B. Olsen, D. D. Maynes, D. Slawson, and K. Ho, "Comparison and equating of paper-administered, computer-administered and computerized adaptive tests of achievement," presented at the Annu. Meet. Amer. Educ. Res. Assoc., San Francisco, CA, 1986.
- [10] J. Mazzeo and A. L. Harvey, "The Equivalence of Scores from Automated and Conventional Educational and Psychological Tests," College Entrance Examination Board, New York, College Board Rep. 88-8, 1988.
- [11] A. C. Bugbee, "The equivalence of paper-and-pencil and computer-based testing," *J. Res. Comput. Educ.*, vol. 28, no. 3, pp. 282–299, 1996.
- [12] S. M. Alessi and S. R. Trollip, *Computer-Based Instruction: Methods and Development*. Englewood Cliffs, NJ: Prentice-Hall, 1991, pp. 205–243.
- [13] L. C. Jacobs and C. I. Chase, *Developing and Using Tests Effectively: A Guide for Faculty*. San Francisco, CA: Jossey-Bass, 1992, pp. 168–177.
- [14] R. J. Shavelson, G. P. Baxter, and J. Pine, "Performance assessments: Political rhetoric and measurement reality," *Educ. Res.*, vol. 21, no. 4, pp. 22–27, 1992.
- [15] F. M. Bernt, A. C. Bugbee, and R. D. Arceo, "Factors influencing student resistance to computer administered testing," *J. Res. Comput. Educ.*, vol. 22, no. 3, pp. 265–275, 1990.
- [16] K. Rasmussen, P. Northrup, and R. Lee, "Implementing Web-based instruction," in *Web-Based Instruction*, B. H. Khan, Ed. Englewood Cliffs, NJ: Educational Technology, 1997, pp. 341–346.
- [17] B. H. Khan, "Web-based instruction (WBI): What is it and why is it?," in *Web-Based Instruction*, B. H. Khan, Ed. Englewood Cliffs, NJ: Educational Technology, 1997, pp. 5–18.
- [18] J. Ravitz, "Evaluating learning networks: A special challenge for Web-based instruction," in *Web-Based Instruction*, B. H. Khan, Ed. Englewood Cliffs, NJ: Educational Technology, 1997, pp. 361–368.
- [19] T. Ager, "Online placement testing in mathematics and chemistry," *J. Comput.-Based Instruc.*, vol. 20, no. 2, pp. 52–57, 1993.
- [20] M. Sweeney, M. Maguire, and B. Shackel, "Evaluating user-computer interaction: A framework," *Int. J. Man-Mach. Stud.*, vol. 38, pp. 689–711, 1993.
- [21] Quiz Wiz [Online]. Available: <http://www.ccts-ent.com/qwiz/index.html>
- [22] Y. D. Lin, C. Chou, Y. C. Lai, and W. C. Wu, "WebCAT*—A Web-centric, multi-server, computer-assisted testing system," *Int. J. Educ. Telecommun.*, vol. 5, no. 3, pp. 171–192, 1999.
- [23] C. Chou, "Developing CLUE: A formative evaluation system for computer network learning courseware," *J. Int. Learn. Res.*, vol. 10, no. 2, pp. 179–193, 1998.
- [24] J. E. Pitkow and M. M. Recker, "Using the Web as a survey tool: Results from the second WWW user survey," *Comput. Networks ISDN Syst.*, vol. 27, no. 6, pp. 809–822, 1995.

Chien Chou (M'95–A'96) received the B.A. degree in history from National Taiwan University, Taipei, Taiwan, R.O.C., in 1984, the M.S. Ed. degree in instructional systems technology from Indiana University, Bloomington, in 1986, and the Ph.D. degree in instructional design and technology from The Ohio State University, Columbus, in 1990.

From 1989 to 1990, she was with Battelle Memorial Institute, Columbus, where she participated in research projects on human factors and training systems. From 1990 to 1993, she was with RCS Research Consulting Services Company, Buffalo, NY, where she developed multimedia industrial training projects. Since 1993, she has been with National Chiao Tung University, Taiwan, where she is currently a Professor in the Institute of Communication Studies. She is engaged in research and teaching in the areas of multimedia message design and evaluation, instructional technology, and distance learning.