DOCUMENT RESUME

ED 391 471 IR 017 619

AUTHOR Glowacki, Margaret L.; And Others
TITLE Developing Computerized Tests for Classroom Teachers: A Pilot Study.
PUB DATE [9 Nov 95]
NOTE 13p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Biloxi, MS, November 8-10, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Computer Assisted Testing; Computer Uses in Education; Courses; Educational Technology; Higher Education; Pilot Projects; *Student Attitudes; Student Surveys; Tables (Data); *Test Format; Undergraduate Students
IDENTIFIERS Paper and Pencil Tests

ABSTRACT

Two types of computerized testing have been defined: (1) computer-based testing, using a computer to administer conventional tests in which all examinees take the same set of items; and (2) adaptive tests, in which items are selected for administration by the computer, based on examinee's previous responses. This paper discusses an option for classroom teachers that is easier to develop than a computerized adaptive test, but more secure and sophisticated than a computer-based test. The process of developing and pilot testing the computer-administered test and the results of a survey of student reactions are described. Subjects for the study consisted of 108 undergraduates taking summer educational technology courses in computer applications at a Southern university. Identical items were used for paper-and-pencil and computerized tests. No significant differences were found for either administration. Student responses indicated that: all of the students had familiarity with computers; 94% had no problems understanding the test directions; 53% initially experienced anxiety about taking the test on a computer; 89% indicated that the computer test was as fair as a paper test; and 61% indicated a preference for the computer test, while 19% indicated that both methods worked equally well. Four tables depict results for computerized tests versus paper-and-pencil tests; descriptive data for both kinds of tests; students' yes/no responses to the attitude survey; and examinee's comments regarding computerized testing. (AEF)

Developing Computerized Tests for Classroom Teachers: A Pilot Study

by

Margaret L. Glowacki, Anna C. McFadden, and Barrie Jo Price
The University of Alabama

Paper presented at the annual meeting of the Mid-South Educational Research Association, November 8-10, 1995, Biloxi, Mississippi.

Introduction

In most public schools and universities, providing routine student examinations involves considerable time and difficulty in developing, copying, administering, and scoring examinations. Some educational programs use computerized scoring of scannable answer sheets, but this requires a scanner, computer, software, scannable answer sheets, someone to run and maintain the equipment, and students are required to remember to bring a No. 2 pencil to bubble the answer sheets. Tests administered using scannable answer sheets may be cost effective for large organizations, but generally this alternative is impractical for public school systems where tests are administered to small groups of students, and each teacher uses a different test. However, computers and the development of computer-based testing offer an alternative to classroom teachers.

The use of computerized testing in education is becoming more commonplace due to the availability of hardware and software. Wise and Plake (1990) discussed the advantages of using computer-administered testing in educational settings. The first of these is time. Research has indicated a reduction in testing time over conventional tests when using computer-administered testing (English, Reckase, & Patience, 1977; Olsen, Maynes, Slawson, & Ho, 1989). Other advantages include feedback regarding scores and incorrect responses, the ability to gather additional information about the testing session, greater standardization, and the ability to administer new item types not possible with conventional tests. Previously, one of the obstacles to using computer-based testing in classrooms was the teachers' lack of time and programming skills. With the use of authoring packages and already developed models, teachers will not need to learn programming skills, and will actually save time.

Wise and Plake (1989) differentiated between two types of computerized testing: computer-based testing and computerized-adaptive testing. They defined computer-based testing as using a computer to administer conventional tests, in which all examinees take the same set of items, while adaptive tests are defined as tests in which items are selected for administration by the computer, based on examinee's previous responses. This paper discusses an option for classroom teachers that is easier to develop than a computerized adaptive test, but more secure and sophisticated than a computer-based test. Using *Authorware Professional* by Macromedia, an authoring package, tests were developed that are completely administered and scored using a computer, and that can be programmed to randomly administer subsets of items from an item pool. This paper describes the process of developing and pilot testing the computer-administered tests and the results of a survey used to obtain student reactions.

Method

Subjects

The subjects for this study consisted of a sample of 108 undergraduates taking summer educational technology courses in computer applications at a Southern university. The students were from several programs across the campus including teacher education, business, arts and sciences, communications, marketing, and advertising. All of the students taking the course for the summer session were involved in the study, although

two students received incompletes and did not take the second test. Because of this, their scores were dropped from the analyses for the first test, although their attitude data were included.

Instruments

There were three instruments developed for this study: two tests consisting of 30 true-false items, with each test covering different units of study for an undergraduate course in educational computer technology applications, and a survey instrument to elicit responses from students about their reactions to the computerized tests.

The items for the course tests were developed by the professors teaching the course. The identical items were then used for both the paper-and-pencil (paper) and computerized tests. Three equivalent forms of the paper test were created, all containing the same items, but with the items appearing in a different order on each form. *Authorware Professional*, an authoring package for computer applications development, was used to create the computerized tests. *Authorware Professional* is an object-oriented authoring program that allows the development of interactive applications. *Authorware Professional* is available for both Macintosh and DOS computers and allows cross-platform development, which means any test created using a DOS computer can be converted to Macintosh format and vice versa. Applications created in *Authorware Professional* are "packaged" with run-time files, which allows the creation of executable files that do not require having *Authorware Professional* installed on a computer to run the applications. The computerized tests also consisted of identical items, but the items were administered in random order to all students, which meant each student answered the items in a different order. Other items on the computerized tests required students to type their names and student identification numbers, and as a security measure, a set of items was used to collect personal information already available from course enrollment forms to be certain the individuals taking the exam were who they said they were. The items were presented one at a time on the screen. When the students responded to an item, the next screen told them what they had answered and then gave them a chance to go back and change the answer immediately if they wished. A chance to review correct answers to any missed items was offered to all computer examinees.

When the computerized tests are administered, the answer to each item is recorded as well as the examinees' total scores, whether or not they chose to review missed items, and the time needed to take the test. All of the data collected is written to a text file that can then be imported into any spreadsheet or statistical program to analyze the data. Immediately upon finishing the test, the examinees are given their score and then asked whether they wish to review the items they missed.

The survey developed to elicit students' reactions consisted of 18 items requiring yes/no, Likert scale, and open-ended responses. The survey also was administered on the computer immediately following the test. The students answered the survey items once, when they took the computerized test.

Procedure

The subjects were randomly selected to take the first test in either the computer format (Computer Test #1) or the paper-and-pencil format (Paper #1) so that one-half of the students took the computer test and the other half took the traditional test. For the second test, the groups were switched (Paper #2 and Computer Test #2). After each computer administration, students were asked to complete the survey soliciting their reactions to the computerized test.

Data Collection and Analysis

The tests were administered on DOS computers during regular class periods with one-half of the students taking the paper-and-pencil test in one computer lab and the other one-half of the students taking the computerized test in a different computer lab. The data from both tests were imported or typed into Microsoft *Excel*, a spreadsheet program; *t*-tests were performed to compare the two groups; and descriptive data were examined. The survey data also were imported, and frequencies and percentages calculated using Microsoft *Excel*.

Results

Computer Tests

Table 1 shows the results of the *t*-tests conducted on the two test administrations. There were no significant differences found for either administration. Additional statistical information is provided in Table 2. The means for the first exam were 25.20 for the computerized test and 25.38 for the paper-and-pencil test. The means for the second test were 22.62 for the computerized test and 22.78 for the paper-and-pencil test. The medians, modes, and variances were essentially identical.

Table 1
t-Test Results for Computerized Tests vs. Paper-and-Pencil Tests

n=106	*df*	*t*	*p*
Test #1	104	-0.3850	0.3501
Test #2	104	0.2816	0.3887

$p < .05$

Table 2
Descriptive Data for Computerized and Paper-and-Pencil Tests

	Computer Test 1	Paper-and-Pencil Test 1	Computer Test 2	Paper-and-Pencil Test 2
N	51	55	55	51
Mean	25.20	25.38	22.78	22.68
Standard Error	0.36	0.33	0.46	0.37
Median	26	26	23	23
Mode	26	26	25	24
Standard Deviation	2.55	2.40	3.28	2.75
Variance	6.52	5.76	10.73	7.57
Range	14	11	15	15
Minimum	14	19	14	15
Maximum	28	30	29	28

Attitude Survey

The results of the yes/no and Likert-scale items on the attitude survey are reported in Table 3. Responses indicated that all of the students had familiarity with computers and all had taken at least one prior computer course, with 79% previously having taken two or more computer classes; 39% owned their own computers; 82% reported using computers for class assignments in other classes; 62% reported using computers daily or 1-2 times per week; and only 14% used a computer less than once per month. Ninety-one percent (91%) of the students reported feeling comfortable using a computer. Only 22% of the students had previously taken a computerized test.

When asked if they had trouble understanding the test directions, 94% reported they had no trouble and 6% reported that the instructions could be improved. No students reported that the instructions were unclear or difficult to understand. A high percentage of students (94%) reported they had no difficulty reading the questions on the screen. One student reported that the print was difficult to see due to the font used (Courier), one student reported the print was difficult to read because of its location, and four students (4%) reported that the print was difficult to read for other reasons.

Nearly all the students (99%) reported they had no problems with the computer while taking the test. One student reported that he or she had a minor problem, but that it did not interfere with taking the test. When asked to identify the problem, the student stated that he or she had typed his or her last name incorrectly and could not go back and change it. That item has since been revised and the examinees are asked to check their name to be sure it is correct, and if not, to return to that screen and change it.

There was some anxiety about the test, with 53% reporting that they experienced more anxiety about taking the test on a computer than they would have in a paper form, but this anxiety disappeared after the first administration. After completing the test, students were given their scores and then were given the option of

reviewing the correct answers to any items they missed. Eighty-two percent of the students reported they liked this feature and did review the items; 12% liked the feature but did not review their incorrect responses. When asked, "Do you like being able to receive your score immediately after finishing the exam?," 96% reported yes and the remaining 4% said it makes no difference.

Although examinees are unable to skip items or go back and review the entire test, they are given the option of going back and changing the answer to an item before moving on the next item, in case the examinee clicked the wrong answer or pressed the wrong key. Eighty-two percent of the examinees said they had no problems changing answers to questions; 14% responded that there were a few answers they wished to change but could not; and 4% said they had serious problems, that there were answers they wished to change and could not. These students were handled individually by the instructor. Ninety-four percent of the students reported they would like to be able to go back and review all questions and be able to change answers, while the other 6% responded that even given the option, they would not have gone back to change answers. Some students reported verbally to the instructor that had they had the opportunity to go back and change answers, they would have missed more items on the test, that their initial responses were correct.

When asked if the computer test was as fair as a paper test, 89% indicated in the affirmative. Interestingly, 92% reported that computerized test took less time than a paper test. However, the average time for the first test was 25 minutes, 13 seconds, with a range of 76 minutes and a maximum time of 47 minutes. The second test was not much different, with 21 minutes, 57 seconds, and a minimum time of 5 minutes and 45 minutes for the maximum time. In classroom administrations using paper tests, the range is about the same for these two tests.

Eighty percent responded that if given the opportunity to take the next test using a computer, they would choose to do so. When asked which methods they would prefer for testing: computer instead of paper, paper instead of computer, or no preference, 61% indicated a preference for the computer test and 19% indicated that both methods work equally well. For the open-ended items, 98 students registered remarks about the test. The overwhelming majority of responses were positive, most of them reflecting that it was "quick" or "easy" or even "fun," or that the test was good because it gave them immediate performance results. Table 4 lists some of these responses.

Discussion

The tests used were one type of objective test that is commonly used by classroom teachers for examining student performance. Depending upon local circumstances, the development of tests can be a cumbersome part of a teacher's role. In this case, 106 subjects completed two tests during the pilot study. Even in short tests of this kind, a minimum of 6,360 items needed to be scored. If scored by hand, this would probably involve a minimum of one to two hours to score the tests. Another half hour might be devoted to recording the grades, and then the papers would have to be organized, stored, and carried to class to show students their results. Assuming

two hours labor per test, this is "lost" time that was saved by using the computerized format. The majority of time spent on the computerized tests is in development: writing the items (which is true of both formats) and then putting them into the computer program. But once the initial test has been created, it is simple to revise, add, and delete items. There is no time spent in copying tests for each administration, grading all of the tests, or recording grades. This is all automated.

Other benefits of computer testing include a reduction in testing time (English, Reckase, & Patience, 1977; Olsen, Maynes, Slawson, & Ho, 1986), security, flexibility (tests can be administered at almost any time), immediate feedback, immediate scoring and reporting of scores to examinees, ease of taking the test for physically disabled students who cannot write, and additional information that can be gathered about the testing session. The benefits of flexibility and security are due to the fact that the tests can be hidden on a computer or kept on a floppy disk until time to administer the test, so there is less chance of the test being seen before administration. *Authorware Professional* also allows random administration of items, so even if an entire class is taking the test at the same time, no two students will likely be taking the same items at the same time. This means students can sit right next to each other and be unable to look at each others' answers. Because of the flexibility and security, even teachers with one computer in the classroom could use computer-administered tests.

One area of concern in computer-administered testing is the ability to skip items, review already answered items, and the ability to change answers. As previously discussed, when responding to the attitude survey item asking the examinees to respond to the fact that they could not go back and review the entire test and change answers, 94% of the examinees responded that they would like to be able to do this. Much consideration was given to this feature at the time the tests were being developed, but the programming involved in allowing this option would be much more complex, the test files would be much larger, and it would be more difficult for classroom teachers to gain the time and expertise to develop the tests. Also, because independence of items is sometimes difficult to achieve, not being able to return to previous items made it less likely that students could figure out answers to an item they didn't know, from another item. This was witnessed during one administration. A student did not know an item, but commented when taking a later item, that if she could go back, she could now answer the first item. Wise and Plake (1990) recommended that because of the more complex programs and longer testing time these features would involve, unless research finds that not including the ability to skip items, review already answered items, and change answers affects the test scores of a meaningful proportion of examinees, it is not necessary to provide these features. Harvey (1987, as cited in Wise & Plake, 1990) conducted a study in which he compared a computer-based test that included these three features with one that did not. He found no significant differences, however, his study must be interpreted cautiously because the subjects were college students participating in the study for research credit. In the current study, although students taking the pencil and paper tests were allowed to review and change their answers and the

students taking the computerized exams were not, there was still no significant difference between the testing methods.

As with much of the previous research, this study did not find any significant differences between the computer-based testing and conventional testing. Bunderson, Inouye, and Olsen (1989) examined and summarized the findings of research conducted on the issue of equivalence between computer-based and conventional testing. They found reliabilities of computer-based and conventional testing to be very similar, and that when differences were found, they were generally insignificant, although the computer-based testing generally produced lower scores than conventional testing. This also was true for the current study. Bunderson et al. (1989) concluded that the differences held little practical significance. However, Wise, Barnes, Harvey, and Plake (1989), suggested that these differences may be due to only a small percentage of the examinees whose scores were significantly affected by the testing. Bugbee and Bernt (1990) conducted studies involving students taking entrance examinations for The American College. They found that the students taking the computer-administered tests tend to do as well as, if not somewhat better than those taking paper-and-pencil tests. The results of this study could be due to the fact that the student body consisted of insurance and financial agents, most of who use computers on a regular basis. Bugbee and Bernt concluded the results also could be because computerized tests are less prone to random error due to misrecording answers. Bugbee and Bernt also collected attitude data in which the students indicated two main reasons for preferring the computer-administered tests: immediate knowledge of their grades, and the fact that they could schedule the test whenever they were ready.

The computerized tests have now been administered in the educational technology courses used in this study for several semesters. The students continue to make positive comments about the testing format, and none have expressed the desire to return to paper-and-pencil tests. The data from each test is analyzed using item difficulty and discrimination indices and items are dropped or revised accordingly. Since several tests are administered per course, test grades for each students are examined to see if they are consistent. If a large discrepancy is discovered, it generally has been found that the student did not study for one of the exams. Because the items are administered from a pool of items, and each student may take a different set of items, if the discrepancy is thought to be due to the selection of items the student received then the test is re-administered to that student.

The authors have continued to refine the tests and the testing process. Other types of items such as multiple choice and short answer have been incorporated into some of the tests. The tests also have now been placed on a local area network which means that the students can take the tests any time they are ready rather than taking them as a group. This also makes administering make-up tests much easier. Most of the students taking the educational computer technology courses are pre-service or inservice teachers and many have decided to develop the computerized tests for their own classrooms.

This study confirms the findings of many other studies regarding the equivalence of computer-based and conventional testing. Because other studies have contradicted these results, test developers must continue to perform equivalence studies on tests as they are developed. Other concerns that must continue to be explored in computer-based testing are that students perform in similar ways on the computer-based and conventional tests, issues regarding the presentation of the items on the screen, feedback, and the ability to review and change test items.

References

Bugbee, A. C., & Bernt, F. M. (1990). Testing by computer: Findings in six years of use 1982-1988. Journal of *Research on Computing in Education, 23*(1), 87-100.

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 367-408). New York: Macmillan.

English, R. A., Reckase, M. D., & Patience, W. M. (1977). Application of tailored testing to achievement measurement. *Behavior Research Methods and Instrumentation, 9*, 158-161.

Harvey, A. L. (1987). *Differences in response behavior for high and low scorers as a function of item presentation on a computer assisted test.* Unpublished doctoral dissertation, University of Nebraska, Lincoln.

Olsen, J. B., Maynes, D. D., Slawson, D., & Ho, K. (1989). Comparisons of paper-administered, computer-administered, and computerized adaptive achievement tests. *Journal of Educational Computing Research, 5*(3), 311-326.

Wise, S. L., & Plake, B. S. (1990). Computer-based testing in higher education. *Measurement and Evaluation in Counseling and Development, 23*, 3-10.

Wise, S. L., & Plake, B. S. (1989). Research on the effects of administering tests via computers. *Educational Measurement: Issues and Practice, 8*(3), 5-10.

Wise, S. L., Barnes, L. B., Harvey, A., & Plake, B. S. (1989). The effects of computer anxiety and computer experience on the computer-based achievement test performance of college students. *Applied Measurement in Education, 2*, 235-241.

Table 3

Responses to Likert and Yes/No Items on the Attitude Survey of Examinees' Responses (N = 108)

Item	Frequencies	Percentages
1. How many computer courses have you taken?		
One class	22	21%
Two classes	41	39%
Three classes	32	30%
Four classes	8	7%
Five classes	1	1%
Six classes	1	1%
Seven classes	1	1%
2. Do you own a computer?		
Yes	42	39%
No	66	61%
3. Do you use the computer for work other than assignments for this computer class?		
Yes	89	82%
No	19	18%
4. How often do you use a computer for things other than work for this class?		
Daily	22	20%
1-2 Times Per Week	45	42%
1-2 Times Per Month	26	24%
Less than 1 Time Per Month	15	14%
5. Have you taken computerized exams before?		
Yes	24	22%
No	84	78%
6. Do you feel comfortable using a computer?		
Yes	98	91%
No	10	9%
7. Did you understand the instructions for the exam?		
Yes, I had no trouble understanding them.	102	94%
I understood them, although they could be improved.	6	6%
The instructions were unclear and difficult to understand in some parts.	0	0%
No, I did not understand the instructions.	0	0%
8. Was it difficult to read the questions on the screen?		
No, I had no difficulty.	102	94%
Yes, the print was difficult to read mainly because of the type of font.	1	1%
Yes, the print was difficult to read mainly because of its size.	0	0%
Yes, the print was difficult to read mainly because of its location.	1	1%
Yes, the print was difficult to read for other reasons.	4	4%
9. I was more anxious about taking the test on a computer than I would have been if I were taking it on paper.		
Yes	57	53%
No	51	47%

Table 3 continued

10. Did you like having the opportunity to review the questions you answered incorrectly?		
Yes, and I took the opportunity to review those items I answered incorrectly.	88	82%
Yes, I liked having the opportunity, although I did not review.	13	12%
Not for myself, although it is good to have the opportunity for others to review.	7	7%
No, I would prefer that the review not be part of the exam.	0	0%
11. Do you like being able to receive your score immediately after finishing the exam?		
Yes, I do.	104	96%
No, I don't.	0	0%
It makes no difference.	4	4%
12. Did you have any problems with the computer while taking the exam?		
No, I had no computer problems.	107	99%
Yes, but they were minor problems and did not interfere with taking the exam.	1	1%
Yes, and the problems interfered with taking the exam.	0	0%
Types of Problems Encountered		
I TYPED MY LAST NAME IN WRONG AND COULD NOT GO BACK.		
13. Did you have any problems changing your answers to any of the questions?		
No, I was able to change all the answers I wished to.	89	82%
Yes, I had some problems; there were a few answers I wished to change but could not.	15	14%
Yes, I had serious problems; there were many answers I wished to change but could not.	4	4%
14. Would you like to have the opportunity to go back and review or change your answers once you have answered all the questions?		
Yes, I would like to be able to go back and review all of the questions and be able to change the answers if I wished.	102	94%
No, I would not have gone back over the questions anyway.	6	6%
15. The computer exam was as fair as an exam administered on paper.		
Yes	96	89%
No	12	11%
16. The computerized exam took less time than an exam on paper.		
Yes	99	92%
No	9	8%
17. Given the opportunity, I would choose to take the next exam on a computer.		
Yes	86	80%
No	22	20%
18. Which method would you prefer for taking an exam?		
I would prefer taking an exam on the computer rather than on paper.	66	61%
I would prefer taking an exam on paper rather than on a computer.	22	20%
I have no preference, both methods work equally well.	20	19%

Table 4
Examinees' Comments Regarding Computerized Testing

Item: I liked taking the exam on the computer because:

It allowed for a quicker test time. In other words, it made taking the test a whole lot more fun and it was much easier than having to fill out a scan-tron sheet or a blank sheet of paper. This method of testing is fabulous and should be used in every testing situation, except an essay test.

Every time I use the computer, it increases my skills and proficiency level so that I might be able to compete with students from other countries better.

It was easier and less time consuming. The only thing is that I would like to be able to review all the questions again before it is graded.

It was easier and faster. It was more convenient and I knew my scores immediately, I didn't have to wait until the papers were graded. I felt more at ease working on the computer because I'm used to it.

Time is saved by using the computer. Also, common Scantron errors are avoided since there is no bubbling. Plus, as an environmentalist, paper is conserved.

It was fun and it made me feel more at ease. This computer helped me to take my time and not rush my answers. Even though I passed I still would have liked to make a perfect thirty.

Since I work on computers, I feel more comfortable taking the exam on computer than on paper. The computer is not hard to use and I think it is a good way to take a test.

The questions I answered first eased my anxiety about the test. The computer made the test more interesting and fun to take. I liked best that I didn't start right into the test but had a few questions to take my mind off things.

I wasn't as nervous and it wasn't as intimidating as paper exams

It was very fast and efficient. It was more relaxing than having a bunch of questions and answers staring you in the face. Computerized testing lets you only think about the question you're working on. Your mind can not wonder about another question.

It allowed me to work at my own pace to start and finish the exam. I was able to work at my own pace. I felt that the test was very well explained and provided no confusion at all. I would like to take all of my test on a computer.

The format and screens were interesting, which is very unusual for an exam. However, I felt uncomfortable with the fact that I could not go back and review my answers and perhaps make changes. I really liked getting my score as soon as I finished the exam and being able to review my incorrect answers.

I did not have to bring a #2 pencil. Fast test grade. I liked it.

It is easier to read. It helps me to see one question at a time so that I am not overwhelmed by the whole exam. I only have to worry with one question at a time. I had no problems with it.

It gave me my grade right away and it told me why the ones I missed were incorrect.

It was so easy compared to doing it on paper. To be truthful it is a lot less hassle than a paper test.