DOCUMENT RESUME

ED 384 679 TM 024 001

AUTHOR	Powers, Donald E.; O'Neill, Kathleen
TITLE	Inexperienced and Anxious Computer Users: Coping with a Computer-Administered Test of Academic Skills. The Praxis Series: Professional Assessments for Beginning Teachers.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-92-75
PUB DATE	Oct 92
NOTE	53p.
PUB TYPE	Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Beginning Teachers; *College Students; Computer Anxiety; *Computer Assisted Testing; Computer Attitudes; Computer Literacy; *Experience; Higher Education; *Teacher Evaluation; Test Construction; Testing Problems
IDENTIFIERS	Academic Skills; *Familiarization; *Praxis Series

ABSTRACT

The objective of this study was to assess the degree to which the mode of administration of the computer-based Academic Skills Assessments of the "Praxis Series: Professional Assessments for Beginning Teachers" contributes to performance differences among test takers. Inexperienced or anxious computer users (Praxis pilot sample of 446 college students and graduates and study sample of 145) were recruited to take the assessments. The degree to which test design and test familiarization procedures effectively minimized variation due to comfort and familiarity with computers was examined from three perspectives: (1) the extent to which the availability of a personal, information-providing test center supervisor influenced test performances, beyond the help provided by a computerized test familiarization tutorial; (2) the effect of within-test practice on later performance on a subsequent section of the test; and (3) the relationship of computer-based test performance to attitudes toward computers and experience in using them. The conclusion was that performance on the tests is not unduly affected by computer administration. Eleven tables present study results. Two appendixes, with five more tables, present pilot and study samples' responses to computer attitude scale items and supplemental information. (Contains 38 references.) (Author/SLD)

THE PRAXIS SERIES
Professional Assessments for Beginning Teachers™

Inexperienced and Anxious Computer Users: Coping with a Computer-Administered Test of Academic Skills

Donald E. Powers

Kathleen O'Neill

October 1992

RR. 92-75

ETS Educational Testing Service

Inexperienced and Anxious Computer Users:

Coping with a Computer-Administered Test of Academic Skills

Donald E. Powers

Kathleen O'Neill

October 1992

THE PRAXIS SERIES: PROFESSIONAL ASSESSMENTS FOR BEGINNING TEACHERS™

Acknowledgments

The authors extend sincere thanks to the following people for their significant contributions to this study:

Nancy Cole, for suggesting a study design that incorporated supervisor help as a factor

Laura Jerry, for processing and analyzing the data

Charlie Lewis, for advising us on data analysis

Brenda Loyd, for permission to use the Computer Attitude Scale

Paul Ramsey, for initiating the study

Clyde Reese, for advising us on the design of the study

Lora Vogel, for researching and identifying information about measuring attitudes toward computers

Lou Woodruff and Charlene Canny, for recruiting study participants and arranging the data collection

Ruth Yoder, for painstakingly producing the tables and text of the report

Craig Mills, Phil Oltman, and Paul Ramsey for providing helpful reviews of an earlier draft of the report.

Abstract

The objective of this study was to assess the degree to which the mode of administration of the computer-based Academic Skills Assessments of The Praxis Series: Professional Assessments for Beginning Teachers™ contributes to performance differences among test takers. To make this determination, inexperienced or anxious computer users were recruited to take the assessments. The degree to which test design and test familiarization procedures effectively minimized variation due to comfort and familiarity with computers was examined from three perspectives:

1. the extent to which the availability of a personal, information-providing test center supervisor influenced test performances, beyond the help provided by a computerized test familiarization tutorial,
2. the effect of within-test practice on later performance on a subsequent section of the test, and
3. the relationship of computer-based test performance to attitudes toward computers and experience in using them.

The conclusion was that performance on the tests is not unduly affected by computer administration.

Inexperienced and Anxious Computer Users:

Coping with a Computer-Administered Test of Academic Skills

Computer-based testing affords a number of measurement opportunities that are not possible, or at least not easily implemented, with paper-and-pencil technology. Among the prospects that have to some degree already been realized are new question formats, alternative models of measurement (e.g., adaptive testing), improvements in test administration, immediate feedback to test takers, and more efficient assessment in terms of the kind and amount of information that can be gathered in a given period of time (Educational Testing Service, 1992; Green, 1988; Wise & Plake, 1990). In the future, as computer technologies continue to develop, even more novel and intelligent uses of computers are likely to arise (Bunderson, Inouye, & Olsen, 1989).

Despite the many potential advantages of computer-based testing, there is, as with any innovation, a need to address the possibility of unwanted side effects, including any inadvertent change in the meaning of test scores. Green (1988) has discussed a number of differences between computer-based and conventional tests that may affect the interpretation of test scores obtained in each of these modes.

Arguably the most serious of the potential unintended consequences of computer-based testing is the introduction of construct irrelevant factors that may disadvantage some groups of test takers. Inequity may arise in the context of computer-based assessment to the extent that test taking involves procedures with which every test taker is not equally comfortable or facile (U.S. Congress, Office of Technology Assessment, 1992). Of relevance here is that, according to one estimate (Weil, Rosen, & Sears, 1987) for instance, "a sizable minority" (perhaps as many as one of every three adults) suffers aversive reactions to computer-related technology, ranging from "mild discomfort" to "severe debilitation." There is also ample evidence to suggest that not everyone coming through the U.S. educational system has had equal access to computers or is equally skilled in the use of them (Martinez & Mead, 1988). Furthermore, attitudes toward computers and competence in their use may be related to both gender and ethnicity (Dambrot, Watkins-Malek, Silling, Marchall, & Garver, 1985; Martinez & Mead, 1988; Wilder, Mackie, & Cooper, 1985), although these differences may depend on exactly how computers are used (Lockheed, 1985).

One approach to equity involves examining the equivalence of computerized and paper-and-pencil versions of a test.[1] In accordance with APA guidelines for computer-based tests (American Psychological Association, 1986), several studies have in fact investigated the equivalence of scores from automated and conventional paper-and-pencil test versions. Mazzeo and Harvey's (1988) summary of these efforts suggests that under some conditions computer-based tests are more or less equivalent to their paper-and-pencil counterparts. A very thorough recent comparison of computer-based and paper-and-pencil versions of the GRE General Test, for example, (Schaeffer, Reese, Steffen, McKinley, & Mills, in press) showed an extremely tight equivalence between scores based on the two versions. However, under other circumstances, for example when speed is a factor in performance, scores from automated and conventional tests do <u>not</u> appear to be directly comparable.

While applicable generally, the concept of equivalence also pertains to specific subgroups of examinees, including perhaps those whose attitudes toward or experiences with computers are not commensurate with those of test takers in general. There has been some conjecture (e.g.,

[1]It should be noted that, though relevant, the concept of equivalence is not entirely germane to the computer-based Academic Skills Assessments of The Praxis Series: Professional Assessments for Beginning Teachers™, the test studied here, as this test has no comparable paper-and-pencil counterpart. In contrast, equivalence has been a central issue in converting the existing GRE General Test, for example, to a linear computerized test consisting of the same items used on the paper-and-pencil version of the test.

Wise, Harvey, & Plake, 1989) about the involvement of certain individual differences in test performance differentials resulting from the mode of test administration. Although relatively few studies have examined the role of such traits as attitudes and experience as moderators of test performance on computer-based and paper-and-pencil tests, there have been some exceptions.

Lee, Moreno, and Sympson (1986) hypothesized that anxiety levels may have run higher for the computer-based test than for the paper-and-pencil test used in their study, thus explaining the higher average scores on the paper-and-pencil test. The difference was explained mainly by the lower performance of test takers who had no previous experience with computers; among those having had at least some previous experience, there was no significant difference in performance between modes of administration. In a related study, Lee (1986) found that for a computerized test of arithmetic reasoning, much of the variation (62%) was explained by performance on a paper-and-pencil version of the test. A significant, though small portion of variance (2-3%) was explained by previous computer experience (computer courses, use of word processing, computerized games, and jobs requiring computer use). Wise, Barnes, Harvey, and Plake (1989), on the other hand, found that neither feelings of anxiety nor lack of experience had any effect on performance for a computer-based achievement test for students in an introductory statistics course.

The amount of training that may be required to ensure that test performance reflects the construct of interest rather than the mode of testing is, of course, a central issue here. On the basis of her study, Lee (1986) concluded that "...minimal work with computers may be sufficient to prepare a person for computerized testing" (p.732). Johnson and White (1980) also found that little test familiarization was needed (one hour of experience on the computer) to reduce the disadvantage to elderly examinees taking a computerized version of the Wonderlic Personnel Inventory. Weil, Rosen, and Sears (1987), among others, have developed programs that appear effective in reducing "computer phobia" and, presumably, reducing test mode effects. Some, e.g., Loyd and Gressard (1984a), have found that computer experience is positively related to attitudes about computers. Greater experience apparently does not automatically translate to less apprehension, however (Marcoulides, 1990), and may in fact exacerbate computer anxiety in some instances (Rosen, Sears, & Weil, 1987).

The major concern of the study reported here was whether or not the design of the computer-based Academic Skills Assessments of The Praxis Series: Professional Assessments for Beginning Teachers™ and the test familiarization provided for these assessments were sufficient to ensure that test performances accurately reflect the skills being measured. This concern is much the same as for paper-and-pencil tests -- that performance on a test should not be unduly influenced by familiarity with the procedures required to take it. For computer-based tests, these procedures may include not only those needed for paper-and-pencil tests, e.g., when to make informed guesses and how to use time efficiently, but possibly other ones also, for example those having to do with various aspects of computer use (how to scroll, how to use a mouse, etc.). Our aim, therefore, was to implement a primary standard for educational and psychological testing -- to present evidence that ..."a test does not depend heavily on extraneous constructs" (p. 15) (AERA, APA, NCME, 1985), in this case familiarity, comfort, and experience with computers.

Although computer experience and attitudes toward computers are extraneous to the measurement of the constructs of interest here, i.e., the basic academic skills required of beginning teachers, they are not entirely irrelevant to the practice of teaching itself. If, as has been suggested (Molnar, 1978), computer literacy is indeed a "social obligation" and as much a prerequisite for participation in an information society as is reading literacy, then we might reasonably expect prospective teachers to be able to cope with a computer-based test. Teachers, it can be argued, need to exert some control over new technologies, including computers, in order to employ them effectively in the classroom. The assumption is that attitudes and anxiety about computers will naturally affect teachers' inclinations to use them (Scott, Cole, & Engel, 1992), so much so that some have recommended requiring coursework in computers for pre-service teachers (Koohang, 1987). As Kennewell (1992) has suggested,

> ...computer use in schools will never be fully effective if taught only by those enthusiasts with sufficient tenacity to overcome all obstacles to classroom use... (p. 195)

Nonetheless, even if we can reasonably expect beginning teachers to be somewhat familiar with computers, it would seem that performance on a computer-administered test of basic skills ought not to depend heavily on the degree to which test takers differ in this regard. Accordingly, the main objective of the study reported here was to determine the extent to which familiarity with computers, and attitudes toward them, may contribute to test score differences among examinees. A significant contribution would suggest a need either to modify the assessments (or the tutorial materials) to make them "friendlier" to inexperienced computer users, or to provide more extensive test familiarization to ensure that all test takers are equally well versed in the procedures required to take the tests.

To provide a rigorous test of the adequacy of test design and test familiarization, a decision was made to focus on examinees who were either relatively inexperienced in the use of computers or anxious about using them. The rationale for the study design was as follows. Test familiarization and test design can be regarded as sufficient to the extent that the following conditions pertain:

1. Any extra assistance, beyond the test familiarization materials provided to all test takers, does not have a noticeable effect on test performance
2. Previous practice on one section of the test has no appreciable effect on performance on a subsequent section of the test, and
3. Attitudes toward computers and previous experience with them explain relatively little variation in performance on the computer-administered tests, beyond that explained by performance on paper-and-pencil versions of similar tests.

METHOD

Sample Selection

From September 1991 through March 1992, students were recruited at each of four sites where test centers were established for a pilot test of the Academic Skills Assessments of The Praxis Series: Professional Assessments for Beginning Teachers™, a new set of licensing exams for beginning teachers. These centers were located at Arizona State University (Tempe, AZ), Norfolk State University (Norfolk, VA), University of Wisconsin at Platteville, and University of Wisconsin at Stevens Point. Staff at each location were asked to identify students who were either inexperienced with respect to their use of computers or nervous about using them. Recruitment was targeted at students who intended to pursue a career in elementary or secondary education. The following two questions were used to screen volunteers:

1. How often have you used a computer in the last five years? (Some time each week, some time each month, infrequently, never)
2. How apprehensive are you about using computers? (Not at all, somewhat, very)

Students who responded "infrequently" or "never" to question 1, or "somewhat" or "very" to question 2 were invited to participate in the study.

Design

At each site volunteering students, who were paid for their participation, were assigned to either a NO HELP condition or a HELP condition. In the latter, a test center supervisor was available during the test administration to answer questions about computer operations. Special guidelines were developed with respect to test center supervisors' information-providing roles. In short, for students in the HELP condition, supervisors were allowed to answer any questions having to do either with the mechanics of test taking or with computer operations, but not those seeking information about the content of the test. Supervisors were asked to be supportive and accessible without being obtrusive. Examinees in the HELP condition were told that they could ask as many questions as they wished about how to use the computer -- both when they were using the tutorial and when they were answering actual test questions. Students in the NO HELP condition were told that, because the purpose of the study was to assess the effectiveness of explanations about using the computer, they were to rely solely on the computerized instructions. Only in an emergency would the supervisor be available to offer assistance. Because of logistical requirements, students were assigned in sets of four, rather than individually, to either the HELP or the NO HELP condition. Although the assignment was not strictly random, the procedure was thought to produce reasonably comparable groups. (The validity of this assumption is tested later in this report.)

To determine the size of any test practice effect, the computerized tests were administered in two different orders. Every other student received either the reading test first, followed by the mathematics test, or else the mathematics test followed by the reading test.

Instruments

A battery of measures was administered via computer to each participating student. Prior to testing, all students also completed a paper-and-pencil background questionnaire along with a computer attitude survey. The attitude survey, selected after a review of several available inventories, was one developed by Loyd and Gressard (1984b). The validity of this inventory has been studied extensively (see, for example, Bandalos & Benson, 1990, and Dukes, Discenza, & Couger, 1989), and when compared with other similar measures, this survey has been judged to be especially suited to measuring the attitudes of computer novices (Woodrow, 1991). Of the four scales that comprise the measure, only the anxiety, confidence, and liking scales were of interest for the current study. After completing the inventory, students began to work on the computer.

First, each student spent as long as he/she wished on a package of test tutorials that explained the computerized testing procedures. The test tutorials were composed of four sections: how to use a mouse, how to select an answer, how to use the testing tools (e.g., the help and calculator functions), and how to scroll. Students were required to take each tutorial and they could request to repeat any or all of these sections as desired.

After completing the tutorials, each student took the computerized reading and mathematics tests. Each test consisted of 40 questions; the time allotted for the reading test was 80 minutes and for the mathematics test, 60 minutes. The reading test consisted of passages with associated test questions. Passages were drawn from four different content areas: humanities, social science, science and nature, and teacher-related topics. Questions required examinees to recognize or summarize main or supporting ideas, recognize inferences or assumptions, and identify or interpret relevant information. The several question types involved highlighting sentences, inserting new text, completing a timeline or outline, and selecting several correct choices from a list.

The mathematics test consisted primarily of stand-alone test questions. A four-function on-line calculator was available to examinees throughout the test. The content of the test included mathematical relationships, numerical operations, data interpretation, measurement, and quantitative reasoning. A test question may include visual material such as a graph, scale, or histogram. Three kinds of questions were used: a) multiple-choice questions that require examinees to select a single correct answer, b) multiple-choice questions that require examinees to select several correct answers, and c) free response questions, which require examinees to produce a correct answer.

After completing the computerized tests, students were asked, in a set of paper-and-pencil debriefing questions, to give their reactions to the tests, including their opinion of the adequacy of various explanations and the clarity of instructions. Students were also asked to compare various aspects of the computerized testing experience with those of traditional paper-and-pencil testing.

Finally, all participants took a 15-item paper-and-pencil reading test and a 15-item mathematics test. Items were selected from a retired version of the Pre-Professional Skills Test (PPST), a basic skills test that is currently used in the licensing of beginning teachers. Items were selected so as to represent both the range of content and the range of difficulty in the PPST. In addition, the items represented the kinds of reading and quantitative skills that are measured in the computerized tests. Because the PPST contains items that are relatively easy for most test takers, a greater proportion of difficult items was sampled, in order to constitute as reliable a test as possible in the limited amount of time available.

Data Analysis

Several initial checks were made on the quality of the data. This editing resulted in the deletion of 27 records for several different reasons. Some students retook the tests, and the later records of these students were removed from the data file. Other examinees had extensive patterns of missing data either on the background questionnaire or on the tests. For a few, the experimental condition to which the examinee had been assigned could not be determined. For the paper-and-pencil tests, answer response patterns were examined for evidence of random responding. No such evidence was apparent.

In an additional preliminary step (before test performances were examined), the scores on the computer attitude scales were compared for the HELP and NO HELP groups in order to assess the comparability of these quasi-randomly assigned groups. This examination revealed that the groups were significantly different ($p < .05$) on one of the three scales (the confidence scale). Further inspection of the data also revealed that a number of students in the two grouped had quite high scores on both the confidence and the anxiety scales, indicating that these students were probably not very anxious about using computers. Consequently, thirteen additional cases were deleted from the file because they had high values (greater than 3.7 out of a maximum of 4.0) on both of these scales.

Next, reliability estimates (coefficient alpha) and correlations among all variables were computed. The major analyses were hierarchical regression analyses using as dependent variables the scores on each of the computerized tests. Sets of independent variables were added in a stepwise fashion in order to assess their incremental contributions to explaining performance on the computerized measures. The first variables to be added were scores on the paper-and-pencil measures. This order was consistent with our interest in the contribution of the other variables beyond that explained by examinees' performances on traditional, noncomputerized measures of basic skills.

The next variables to be entered as a set were scores from the three computer attitude scales, followed by two dummy-coded variables reflecting examinees' previous experience with computers. (These two experience variables distinguished examinees who had never used a computer before from those who had used one only rarely, and from those who had used one either regularly or routinely.) The attitudinal variables were added before entering the experience variables. but this choice was purely arbitrary.

Next, the two variables reflecting the experimental conditions under our control were added to the equation. The first variable to be entered reflected the order in which the computerized measures were administered. This was followed by a variable reflecting the HELP condition to which the examinee has been assigned. Again, the order of entry of these two variables was largely arbitrary, because each was expected to be uncorrelated with the other and with each of the other explanatory variables. Both the incremental as well as the unique contribution of each set of variables was computed. Finally, product variables reflecting all possible two-way interactions were entered to explore the possibility that some of the design factors might interact with one another to account for additional variation.

RESULTS

Description of the Measures

Each of the 15-item paper-and-pencil tests was of moderate reliability. A Cronbach coefficient alpha of .76 was computed for both the reading and the mathematics tests. The mean proportion correct for reading items was .49, with a range from .29 to .62. For the mathematics test the comparable mean was .46, with a range from .14 to .79.

The reliability estimates were somewhat higher for each of the 40-item computerized tests -- .89 for both reading and mathematics. The mean proportion correct for the reading test was .57, with a range from .21 to .93. For the mathematics test the mean was .61, with a range from .17 to .97. There was no indication that any of the measures was speeded, as nearly every examinee answered every item, and items at the end of each test were generally no more difficult than items appearing earlier in the tests.

Each of the three 10-item computer attitude scales was relatively reliable, with alpha coefficients of .86 for anxiety, .80 for confidence, and .88 for liking. These scales were scored so that higher scores indicated less anxiety, more confidence, and greater liking for computers. Scores have been reported on a 1 to 4 range for each scale, which for each respondent is the total score divided by the number of questions answered. (Individual item responses were scored from 1 = strongly disagree to 4 = strongly agree, except when statements were worded negatively. For these items, the scoring was reversed.)

Description of the Sample

The final study sample of 145 included 10 students from Arizona State University, 63 from Norfolk State, 24 from the University of Wisconsin at Platteville, and 48 from the University of Wisconsin at Stevens Point. Table 1 displays the demographic characteristics of the study sample, along with the characteristics of students who volunteered for the larger pilot test of the Academic Skills Assessments of The Praxis Series. The study sample was predominately female (82%). Nearly all (97%) participants reported that English was their best language. Most (61%) were White, but a significant minority (39%) were Black. A majority (64%) were education majors, who were relatively evenly distributed with respect to year of undergraduate study. The study sample differed significantly from the pilot study sample in the following ways. A slightly greater proportion was female. Slightly more were first-year students and substantially more were Black. Overall grade-point averages were lower, and fewer students were majoring in education.

Because there was no guarantee that recruitment procedures had identified students who were in fact inexperienced or anxious computer users, it was necessary to verify this goal. Table 2 shows the proportions of the study sample that indicated various degrees of experience with computers. As shown, about a fifth of the sample said they had never used a computer, while a slight majority (54%) reported that their use had been rare (only a few times in the previous five weeks), and 25% said they had used computers on a regular basis. Only 1% conveyed that they were routine users of computers. These figures contrast sharply with those given for the larger pilot sample, and even more so for those from a sample of prospective graduate students who participated in a field test of a computer-administered version of the Graduate Record Examinations (GRE) General Test (Schaeffer, Reese, Steffen, McKinley, & Mills, in press). In the latter sample only 1% said they had never used a computer, while more than a third indicated that they were routine computer users. Although the data for the pilot and the GRE samples cannot be regarded as normative in any strict sense, they do suggest that our study sample was indeed relatively inexperienced with regard to computer use.

Table 3 shows that, when compared with the pilot sample, the study sample was also significantly ($p<.001$) more anxious, less confident, and less disposed to liking computers. Between-sample differences on each scale were about .6 standard deviations, a difference that can be regarded as being medium to large (Cohen, 1977). A better feel for the practical significance of this difference can be obtained by comparing group responses to individual items on the scales. (The two samples were significantly different on all ten items in each of the three scales. For 26 of the 30 items the difference was highly significant, $p < .001$). Typical of the differences were the following. Compared with the pilot sample, the study sample more often agreed or strongly agreed with the anxiety scale statement that "Computers make me feel uncomfortable" (42% vs. 24%). The study sample agreed or strongly agreed less often than did the pilot sample with the confidence scale statement that "I have a lot of self-confidence when it comes to working with computers" (39% vs. 64%). They also agreed or strongly agreed less often with the liking scale statement that "Working with computers would be enjoyable and stimulating" (72% vs. 85%). Appendix A contains more detailed information on responses to each attitude scale item for both samples. All in all, these comparisons suggest that we were successful in identifying the target group for the study, i.e., students who were either inexperienced with or anxious about using computers.

Comparison of HELP and NO HELP Samples

As stated earlier, the major experimental manipulation involved the assignment of students to either a HELP or a NO HELP condition. Because, as mentioned above, strict random assignment was not feasible, the possibility existed that the resulting groups might not be precisely comparable. Table 4 shows, however, that, with two exceptions, the assignment procedure did result in groups that were similar with respect to most background characteristics. For some reason that is not readily apparent, however, a somewhat greater proportion of first- and second-year students than upper-class students were apportioned to the HELP condition. In addition, again for no discernible reason, the HELP group contained about twice as many students who reported visual impairments (other than blindness) than did the NO HELP group.

The average performance of the HELP group was slightly higher than that of the NO HELP group on each of the paper-and-pencil and computerized test measures (Table 5), though except for performance on the computer-based mathematics test, these between-group differences were not statistically significant. With respect to performance on each computerized test, differences between groups were commensurate with performance differences on each paper-and-pencil counterpart. There were no significant differences between groups on any of the three attitude scales.

Treatment Implementation

Of the 73 students who were permitted to seek the help of a test supervisor, only 40% reported doing so. Nine students sought information about how to use the mouse, 7 requested help with navigating on the computer, and 13 sought some other information. In accordance with the instructions they received, none of the students reported asking about the content of test questions.

In general then, students in the HELP condition did not make extensive use of the personalized help that was available to them. One possible reason is that they were reluctant to exercise this option. Another possibility, however, is that students found the directions and test familiarization provided by the computer to be sufficient. Test takers' responses to debriefing questions suggest the plausibility of this latter interpretation. For instance, in response to a question about instructions for taking the test, a clear majority of the study sample (73%) judged that the section "contained all the information I wanted." Relatively few believed that the section should have contained more information about either the content of the test (15%), the test taking strategies (20%), or the testing process (12%). Most test takers in the study (70%) reported that the mouse was "very easy" to use; 27% reported having "a little trouble at first, but then easy to use," 1% reported having "trouble for about half the questions," but none had "trouble for most of the questions."

Reactions to the Tutorial

Test takers were also asked about specific features of the test tutorial. Responses are shown in Table 6. For each of the tutorial features listed, about three-fourths (or more) of all study participants felt that the explanations provided were adequate. Some test takers (from 9% to 20%) felt that they already knew the information provided for the various features, and some (from 3% to 12%) suggested that too much information was provided for the features. Few students (from 1% to 4%) thought that the information was either "too little" or else "unclear" on how to use the mouse, how to move between screens, or how to use the "erase" and "calculator" functions. A slightly higher percentage of test takers (14%) reported the need for either more or clearer information about (or more practice with) the "help" and "scroll" functions. In general, test takers in the study sample appeared to be relatively satisfied with most features of the tutorial.

Regression Analyses of Test Performance Data

Correlations among test scores on the paper-and-pencil tests, the computer-administered tests, the computer attitude scales, and examinees' self-reported previous experience with computers are given in Table 7. The correlations among test scores show the expected results. For example, performance on each computer-based test is more highly related to performance on its paper-and-pencil counterpart than to the other paper-and-pencil test. The three attitude scales exhibit strong intercorrelations, particularly the anxiety and confidence scales (.81), suggesting that there is relatively little that is unique to either of these constructs in the study sample. While none of the attitude scales correlated significantly with performance on either of the computer-administered tests, small, statistically significant correlations ranging from -.20 to -.23 were noted between attitude scales and performance on the paper-and-pencil mathematics test. A correlation of -.22 was also found between the anxiety scale and performance on the paper-and-pencil reading test. The reason for these unexpected correlations, which suggest that the greater the anxiety about computers, the higher the performance on paper-and-pencil measures, is not readily apparent.

As stated earlier, amount of computer experience was coded as two dummy variables to reflect responses to the question "How often do you use a personal computer?" A response of "Never" was coded as 0 on both variables, "rarely" was coded as 1 on the first variable and 0 on the second, and "regular" or "routine" use was coded as 0 on the first and 1 on the second. The multiple correlation of these two experience variables with test scores was not significant either for the computer-administered tests or for the paper-and-pencil tests. Experience did correlate significantly (.22 to .32) with each of the computer attitude scales: the more experience, the more positive the attitude.

Results of the hierarchical regression analyses for the computer-based reading and mathematics tests are given in Tables 8 and 9. For each of these computerized tests, the paper-and-pencil counterpart accounted for over half of the variance -- 51% for reading and 58% for mathematics. Also, in each case adding the other paper-and-pencil test to the equation, e.g., the paper-and-pencil mathematics test to the paper-and-pencil reading test to predict performance on the computerized reading test, accounted for an additional small, statistically significant portion of variance -- an additional 3% for the computerized reading test and an additional 4% for the computerized mathematics test. Beyond this, none of the other variables -- attitudes toward computers, computer experience, test order (i.e., the effect of prior practice on a subsequent section of the computer-based test), or being allowed to request help from the test supervisor accounted for a significant amount of variation in performance on the computer-administered reading test. Furthermore, there was no indication that any of these variables interacted in such a way as to explain any additional variation. For the computerized mathematics test, the conclusion is similar, with one exception. The additional 3% of variance explained by adding attitude scores was statistically significant ($p<.05$).

Table 10 displays the regression weights for each explanatory variable for each of the computerized tests. It should be noted that the additional 3% of mathematics test variance is explained by weighting the confidence scale positively (the more confidence, the higher the test score) and the other two attitude scales negatively (the more anxious about and the less liking for computers, the higher the test score).[2]

Although the order of entry of the variables in the regressions was thought to be logically consistent with the nature of the variables and with the objectives of the study, some readers might have preferred a different sequence. For these readers, Table 10 also shows the squared semipartial correlation coefficients, which reflect the unique contribution of explanatory variables when they are considered simultaneously with all other variables. As can be seen, performance on the two paper-and pencil measures uniquely explained about 25% of the variation in scores for the computerized reading test and about 26% of the variation for the computerized mathematics test. With the exception of the attitude scale scores for the mathematics test, no other variable or set of variables contributed more than 1%. For the mathematics test, scores on the confidence scale accounted for 2% of the variance. (When computerized mathematics scores were regressed on only scores from the two paper-and-pencil

[2]Because these opposite-sign weightings may be due to the high correlations among the attitude scales, an additional regression was run, adding only the confidence scale to the paper-and-pencil scores. While the amount of additional variance explained remained the same when only this scale was added, the weight computed for the confidence scale dropped from 4.22 to 2.53.

tests plus scores from the confidence scale, performance on the paper-and-pencil tests uniquely explained 29% of the variance, and confidence scores explained less than 3%.)

Regression Analyses for the Pilot Sample

Because of the unexpected negative correlations in the study sample between test performances and computer attitudes, we were motivated to examine these relationships in the larger, more experienced/less anxious pilot sample. The concern was that variation in the study sample might be less than in the pilot sample, especially regarding attitudes toward computers and experience with them, i.e., the two variables on which the study sample was selected. To this end, two subsamples of pilot examinees were identified that self-reported either SAT scores (N=50) or composite scores from the ACT Assessment (N=141). (Eleven test takers were common to both subsamples.) Tables B.1-B.5 in Appendix B present the results of these analyses. Tables B.1 and B.2 display the correlations among computer-based test scores, attitude scales, computer experience, and depending on the sample, either SAT or ACT scores. (The pilot sample also took a computerized writing test.)

In neither the SAT nor the ACT sample was there a significantly greater range on any of the attitude scales or on the computer experience variable than in the study sample. Table B.1 shows few statistically significant correlations in this small SAT-taking sample. As in the study sample, each computer-based measure correlated higher with its paper-and-pencil counterpart, in this case SAT scores, than with the other measure, and computer experience correlated positively with computer attitudes. None of the attitude scales nor the degree of computer experience correlated significantly with performance on any of the three computer-based measures. The only significant correlation ($p < .05$) was between computer confidence and performance on the mathematics portion of the SAT.

In the ACT sample (Table B.2), many of the same patterns were apparent. Computer experience correlated significantly with computer attitudes, but attitudes were correlated significantly (.20 to .28) only with the computerized mathematics test, not with either the reading or the writing test. Each of the three attitude scales was also correlated significantly with ACT composite score (.24 to .35).

Tables B.3, B.4, and B.5 show for each subsample the results of the hierarchical regression analyses for the computerized reading test (B.3), the computerized mathematics test (B.4), and the computerized writing test (B.5). As in the analyses for the study sample, paper-and-pencil scores (either SAT or ACT scores) were entered first in the regressions, followed by the computer attitude scales, and then computer experience. In each of the six analyses, scores from the paper-and-pencil tests accounted for a statistically significant portion of variance for the computer-based tests. Neither attitudes nor experience accounted for a significant portion of variation, beyond that explained by either SAT or ACT scores, for any of the three computerized tests.

Results of Posttest Debriefing

As indicated earlier, test takers' reactions to various aspects of the test taking experience were solicited upon completion of testing. One question was, "If this had been a real test (one that counted), how anxious would you have been compared with taking a real paper-and-pencil test?" In response, about 45% said they would have been about as anxious on a paper-and-pencil test, 37% said they would have been less anxious, and 18% said they would have been more anxious. When asked to speculate about their performance on a paper-and-pencil test composed of the same questions, a slight majority (54%) believed that they would have done "about the same," while nearly equal percentages felt they would have done "not as well" (22%) or "better" (23%) on a paper-and-pencil test.

The final question posed to test takers was, "If there were a computer test and a paper-and-pencil test with the same questions, which would you prefer to take?" This same question was also asked of all test takers before they took the tests. Table 11 reveals the responses to this question. As the row percentages show, a slight plurality (39%) had no particular preference for either mode of testing before taking the tests, and nearly equal percentages expressed a preference for computer-based testing (32%) and paper-and-pencil testing (29%). Upon completion of testing, exactly half of all test takers said they would prefer a computer-based test, 26% said they preferred a paper-and-pencil test, and 24% had no preference. Of even more interest are the changes exhibited for each pretest preference category. Of those who expressed no preference before taking the tests, about a third (36%) still had no preference after taking them, and the remainder was equally divided in their preference for computer-based and paper-and-pencil tests. The vast majority (85%) of those who initially preferred a computerized test continued to do so after testing. Only 6% subsequently said they would rather take a paper-and-pencil test, and 9% had no preference after having taken the computerized tests. On the other hand, only 38% of those who initially favored a paper-and-pencil test continued to do so after taking the tests. A nearly equal proportion (36%) subsequently preferred a computer-based test, and 26% had no preference for one mode of testing over the other. In general then, the experience of taking the computerized tests resulted in more positive attitudes toward these tests.

SUMMARY AND DISCUSSION

The aim of the study was to determine the degree to which the mode of testing, i.e., computerized administration, contributes to differences among examinees on the Academic Skills Assessments of The Praxis Series: Professional Assessments for Beginning Teachers™, the new teacher licensing exams being developed by ETS. The specific question of interest was whether or not the current design of the test and the method used to familiarize candidates with testing procedures are sufficient to ensure that, as directed by The Standards for Educational and Psychological Testing, "... the test does not depend heavily on extraneous constructs" (p.15). To assess this dependence, a sample of students who were inexperienced with or anxious about using computers was recruited to take the computerized assessments. Test design and test familiarization were to be regarded as sufficient if the following conditions pertained. First, the availability of extra help from a test supervisor, beyond that provided to all test takers in a test familiarization tutorial, would not increase test scores significantly. Second, within-test practice on an earlier section of the computerized examination would not result in better performance on a subsequent section of the test. Third, neither previous experience with nor attitudes about computers would contribute to test score differences above and beyond those explainable from performance on traditional paper-and-pencil tests of similar skills.

To determine the extent to which each of these situations prevailed, half of the study participants were afforded the opportunity to seek extra help from a test center supervisor, who was given special instructions on how to respond to requests from test takers. To assess the effect of practice, half of the sample took the computerized reading test first, and half took the mathematics test first.

Fewer than half of the test takers who were designated as eligible for extra help actually sought assistance from the test supervisor. As responses to debriefing questions revealed, most examinees regarded the explanations provided by the computer to be adequate. Thus, most examinees apparently did not feel the need to request additional assistance. Perhaps not surprisingly then, allowing test takers the opportunity to seek supervisor assistance did not affect performance on either computerized test. Likewise, there was no effect of having taken another test earlier during the test administration. For instance, test takers who took the mathematics test after taking the reading test did not perform any better on the math test than did those who took the test first. The same was true for the reading test. This finding contrasts with the results of several studies of the effects of within-test practice for paper-and-pencil admissions tests, in which prior within-test practice on certain item types has resulted in significantly better performance later in the test (Faggen & McPeek, 1981; Swinton, Wild, & Wallmark, 1983; Wightman, 1981; Wing, 1980).

In contrast, performance on two short paper-and-pencil tests of modest reliability explained over half of the variation in test scores on each of the computer-based measures -- 54% for reading and 62% for mathematics. However, neither knowledge of examinees' previous experience with computers nor their attitudes (anxiety, confidence, or liking) toward computers accounted for any significant additional variation in reading test performance. For the mathematics test, examinees' confidence with computers accounted for an additional 3% of the variation in scores. Although there are no directly comparable data for paper-and-pencil measures, it is probably reasonable to assume that examinees' confidence in taking paper-and-pencil tests may explain at least as much variation for these more traditional tests.

Finally, a post-examination debriefing revealed that examinees were more positive about computer-based testing after completing the test than before taking it. This finding is not entirely unexpected. O'Neill and Kubiak (1991) found similar results for several classifications of examinees in an earlier pilot test of computer-based questions and also for a larger, more recent pilot sample (O'Neill & Kubiak, 1992). Bernt, Bugbee, and Arceo (1990) also cited several studies as evidence that, at all age levels and across a variety of applications, students who use computers report liking them. Post-test reactions also indicated that a majority of test takers in the study believed that they would have performed about equally well on paper-and-pencil and computer-based tests. The remainder of students were equally divided with respect to whether they believed they would have performed better or worse on a paper-and-pencil test than on a computerized test.

In conclusion, three questions were posed about the effects on computer-based test performances of (a) extra-computer assistance from a test supervisor, (b) within-test practice, and (c) previous experience and attitudes toward computers. None of the answers suggested that the effects were large enough to require any major modifications to the tests or to the test familiarization that is available for them. On the basis of these findings, the conclusion is that, with the current test design and familiarization, the mode of test administration does not appear to be a prominent source of variance in test performance for the Academic Skills Assessments of The Praxis Series: Professional Assessments for Beginning Teachers™. This conclusion is reached for a sample of test takers for whom we might expect these factors to exert a greater influence than for test takers in general. Thus, it would seem that performance on a computer-based test does not necessarily depend heavily on examinees' computer skills, as some critics have assumed (see, for example, the assumptions made by the executive director of the National

Center for Fair and Open Testing, as cited by Watkins, 1992). With the technology now available and with careful attention to both test design and test preparation materials, measurement that is highly construct-relevant is a very attainable goal. It will be desirable in the future, however, to continue to monitor computer-based testing for The Praxis Series to determine the impact of changes in its delivery, should any changes be made. Ongoing monitoring should also help to ensure that the results found here generalize to examinees who take these tests in the future.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Psychological Association. (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: American Psychological Association.

Bandalos, D. & Benson, J. (1990). Testing the factor invariance of a computer attitude scale over two grouping conditions. *Educational and Psychological Measurement, 50,* 49-59.

Bernt, F. M., Bugbee, A. C., Jr., & Arceo, R. D. (1990). Factors influencing student resistance to computer administered testing. *Journal of Research on Computing in Education, 23(3),* 265-275.

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

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Dambrot, F. H., Watkins-Malek, M. A., Silling, S. M., Marchall, R. S., & Garver, J. A. (1985). Correlates of sex differences in attitudes toward and involvement with computers. *Journal of Vocational Behavior, 27,* 71-86.

Dukes, R. L., Discenza, R., & Couger, I. D. (1989). Convergent validity of four computer anxiety scales. *Educational and Psychological Measurement, 49,* 195-203.

Educational Testing Service. (1992). *Computer-based testing at ETS 1991-1992*. Princeton, NJ: Educational Testing Service.

Faggen, J. & McPeek, M. (1981, April). *Practice effects for four different item types*. Paper presented at the annual meeting of the National Council on Measurement in Education, Los Angeles, CA.

Green, B. F. (1988). Construct validity of computer-based tests. In H. Wainer & H. I. Braun (Eds.), *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Johnson, D. F., & White, C. B. (1980). Effects of training on computerized performance in the elderly. *Journal of Applied Psychology, 65,* 357-358.

Kennewell, S. (1992). Computing for the terrified. *Computers and Education, 18,* 195-200.

Koohang, A. A. (1987). A study of the attitudes of pre-service teachers toward the use of computers. *Educational Communications and Technology Journal, 35,* 145-149.

Lee, J. A. (1986). The effects of past computer experience on computerized aptitude test performance. *Educational and Psychological Measurement, 46,* 727-733.

Lee, J. A., Moreno, K. E., & Sympson, J. B. (1986). The effects of mode of test administration on test performance. *Educational and Psychological Measurement, 46,* 467-474.

Lockheed, M. E. (1985). Women, girls, and computers: A first look at the evidence. *Sex Roles, 13,* 115-122.

Loyd, B. H, & Gressard, C. (1984a) The effects of sex, age, and computer experience on computer attitudes. *AEDS Journal, 18,* 67-77.

Loyd, B. H., & Gressard, C. P. (1984b). Reliability and factorial validity of computer attitude scales. *Educational and Psychological Measurement, 44,* 501-505.

Marcoulides, G. A. (1990). Affective considerations in the implementation of computer systems. In J. A. Morrell & M. Fleischer (Eds.), *Advances in the Implementation of Computer Systems.* Greenwich, CT: JAI Press, Inc.

Martinez, M. E., & Mead, N. A. (1988). *Computer competence: The First National Assessment.* Princeton, NJ: Educational Testing Service.

Mazzeo, J., & Harvey, A. L. (1988). *The equivalence of scores from automated and conventional educational and psychological tests: A review of the literature* (College Board Report No. 88-8 and ETS RR No. 88-21). New York: College Entrance Examination Board.

Molnar, A. (1978). The next great crisis in American education: Computer literacy. *AEDS Journal, 12,* 11-20.

O'Neill, K., & Kubiak, A. (1991, April). *The relationship between student background and attitudes about a computerized test of basic skills.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

O'Neill, K., & Kubiak, A. (1992, April). *Lessons learned from examinees about computer-based tests: Attitude analyses.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Rosen, L. D., Sears, D. C., & Weil, M. M. (1987). Computerphobia. *Behavior Research Methods, Instruments, & Computers, 19,* 167-179.

Schaeffer, G., Reese, C., Steffen, M., McKinley, R., & Mills, C. (in press). *Field test of a computer-based GRE General Test.* (GRE Board Technical Report 88-8). Princeton, NJ: Educational Testing Service.

Scott, T., Cole, M., & Engel, M. (1992). Computers and education: A cultural constructivist perspective. In G. Grant (Ed.), *Review of Research in Education* (18). Washington, DC: American Educational Research Association.

Swinton, S. S., Wild, C. L., & Wallmark, M. M. (1983). *Investigation of practice effects on item types in the Graduate Record Examinations Aptitude Test* (GRE Board Report No. 80-1cP; also ETS Research Report No. 82-56). Princeton, NJ: Educational Testing Service.

U.S. Congress, Office of Technology Assessment. (1992). *Testing in American schools: Asking the right questions*, OTA-SET-519. Washington, DC: U.S. Government Printing Office.

Watkins, B. T. (1992). Testing service to offer computer version of the Graduate Record Examination. *The Chronicle of Higher Education*, 1 April.

Weil, M. M. , Rosen, L. D., & Sears, D. C. (1987). The computerphobia reduction program: Year 1. Program development and preliminary results. *Behavior Research Methods, Instruments, & Computers, 19,* 180-184.

Wightman, L. E. (1981, April) *GMAT within-test practice effects studies*. Paper presented at the annual meeting of the National Council on Measurement in Education, Los Angeles, CA.

Wilder, G., Mackie, D., & Cooper, J. (1985). Gender and computers: Two surveys of computer-related attitudes. *Sex Roles, 13,* 215-228.

Wing, H. (1980). Practice effects with traditional mental test items. *Applied Psychological Measurement, 4,* 141-155.

Wise, S. L., Barnes, L. B., Harvey, A. L., & Plake, B. S. (1989). Effects of computer anxiety and computer experience on the computer-based achievement test performance of college students. *Applied Measurement in Education, 2,* 235-242.

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

Woodrow, J. E. J. (1991). A comparison of four computer attitude scales. *Journal of Educational Computing Research, 7,* 165-187.

Table 1

Comparison of Demographic Characteristics

of Pilot and Study Samples

Characteristic	Praxis Series Pilot Sample (N=446)	Study Sample (N=145)	χ^2
Sex (% female)	73	82	4.7*
English best language (%)	98	97	
Educational status (%)			
Freshman	11	21	
Sophomore	26	23	9.6*
Junior	30	27	
Senior	24	21	
B.S. or higher	8	8	
Self description (%)			
Black, African or Afro American	5	39	
American Indian, Native American, Alaska Native	2	1	
Asian American, Pacific American	<1	0	111.3**
Caucasian/White	86	61	
Hispanic/Latino/Chicano/Mexican American/Puerto Rican	6	0	
Overall grade point average (%)			
3.5 - 4.0	22	11	
3.0 - 3.5	34	25	24.4**
2.5 - 3.0	30	40	
below 2.5	13	21	
Education major (%)	85	64	55.8**
Career goal (%)			
Elementary teacher	48	39	
Junior high/middle school teacher	9	5	
High school teacher	20	10	49.2**
Administrator/counselor/other school staff	5	8	
Other or undecided	16	37	

$^*p<.05$. $^{**}p<.01$.

Table 2

Percentages of Study Participants Reporting Various Degrees of Computer Use (Compared with Two Other Samples)

Frequency of Use	Computerized GRE Field Test Sample (N=1017)	Praxis Series Pilot Sample (N=446)	Study Sample (N=145)
Never[1]	1	6	20
Rarely (only a few times in the last five weeks)[2]	18	28	54
Regular use (some time each week)	45	56	25
Routine use (almost daily use)	36	10	1

[1]Response choice for the GRE sample was "Never before taking the test today."

[2]Response choice for the GRE sample was "Rarely (only a few times in the last five years)."

$\chi^2(3) = 75.2$, $p<.001$ for Praxis Series pilot sample vs. study sample.

Table 3

Attitudes Toward Computers of Pilot and Study Groups

Computer Attitude Scale		Praxis Series Pilot Sample (N=446)	Study Sample (N=145)	t
Anxiety	Mean	3.29	2.94	6.1***
	SD	.63	.56	
Confidence	Mean	3.20	2.83	7.1***
	SD	.56	.49	
Liking	Mean	3.05	2.71	6.1***
	SD	.58	.62	

Note. Higher scores represent less anxiety, greater confidence, and more liking.

***$p < .001$.

Table 4

Comparison of Students in HELP and NO HELP Conditions

Characteristic	HELP (N=73)	NO HELP (N=72)	χ^2
Sex (% female)	81	83	0.2
English best language (%)	97	96	0.2
Educational status (%)			
Freshman	19	24	
Sophomore	34	11	12.4*
Junior	23	31	
Senior	15	28	
B.S. or higher	8	7	
Self description (%)			
Black, African or Afro American	32	46	
American Indian, Native American, Alaska Native	1	0	
Asian American, Pacific American	0	0	3.9
Caucasian/White	67	54	
Hispanic/Latino/Chicano/Mexican American/Puerto Rican	0	0	
Overall grade point average (%)			
3.5 - 4.0	11	11	
3.0 - 3.5	34	15	8.6
2.5 - 3.0	37	43	
below 2.5	18	28	
Education major (%)	68	60	2.6
Career goal (%)			
Elementary teacher	41	36	
Junior high/middle school teacher	3	7	
High school teacher	14	6	5.4
Administrator/counselor/other school staff	8	7	
Other or undecided	32	43	
Permanent disabilities (%)	40[a]	19[b]	7.1 *
Frequency of computer use (%)			
Never	16	24	
Rarely	58	50	1.7
Regular	25	26	
Routine	1	0	

[a]Of the 40% (30 students) who indicated disabilities, 27 indicated visual impairment (other than blindness), 2 indicated hearing impairments, and 1 indicated a learning disability.

[b]All 19% indicated visual impairments (other than blindness).

*p<.05. **p<.01.

Table 5

Descriptive Statistics for HELP and NO HELP Groups

Variable		HELP (N=73)	NO HELP (N=72)	*t*
Computer reading	Mean	23.5	22.1	1.0
	SD	7.7	8.4	
Computer mathematics	Mean	25.5	22.9	2.0*
	SD	7.3	7.8	
Paper/pencil reading	Mean	7.7	7.1	1.1
	SD	3.4	3.7	
Paper/pencil mathematics	Mean	7.3	6.4	1.7
	SD	3.3	3.3	
Computer anxiety	Mean	2.9	3.0	1.1
	SD	0.6	0.5	
Computer confidence	Mean	2.8	2.9	1.9
	SD	0.5	0.4	
Computer liking	Mean	2.7	2.7	0.7
	SD	0.7	0.6	

*$p < .05$.

Table 6

Study Participants' Reactions to Test Familiarization Tutorial

	Reaction (%)			
Tutorial feature (Explanation of:)	Information already known	Adequate explanation	Too much information	Too little <u>or</u> unclear information[1]
How to use the mouse	13	85	8	2
How to move to the next or previous screen	20	75	12	3
The "erase" function	10	85	7	1
The "help" function	13	73	3	14
The "calculator"	9	85	3	4
How to scroll	13	74	5	14

Note. Respondents were allowed to check all statements that applied.

[1]This category represents the total response to three questionnaire categories: "information is not clear," "too little information," and "too little opportunity to practice."

Table 7

Intercorrelations among Test Scores, Computer Attitude Scales, and Computer Experience

Variable	Mean	S.D.	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Computer reading	22.8	8.1							
(2) Computer mathematics	24.2	7.6	.74						
(3) Paper/pencil reading	7.4	3.5	.71	.63					
(4) Paper/pencil mathematics	6.9	3.3	.58	.76	.61				
(5) Computer anxiety	2.9	0.6	-.07	-.10	-.22	-.23			
(6) Computer confidence	2.8	0.5	-.01	-.00	-.15	-.20	.81		
(7) Computer liking	2.7	0.6	-.04	-.09	-.13	-.22	.59	.68	
(8) Computer experience[1]	0.5, 0.3	0.5, 0.4	.13	.10	.04	.03	.32	.29	.22

Note. With a sample of 145, correlations of .163 are significant at the .05 level. Correlations of .214 are significant at the .01 level, two-tailed test.

[1]Computer experience was coded as two dummy variables. The correlations given here are multiple R's of the two dummy variables with each of the other variables.

Table 8

Results of Hierarchical Regression Analyses

for Computer-Administered Reading Test

Explanatory Variables Added	Cumulative R^2	Increase in R^2	F for Increase in R^2	df
(1) Paper-and-pencil reading	.51	.51	148.6***	1, 143
(2) Paper-and-pencil mathematics	.54	.03	10.7***	1, 142
(3) Computer anxiety, computer confidence, computer liking	.56	.02	1.7	3, 139
(4) Computer experience	.57	.01	2.0	2, 137
(5) Test order (practice)	.58	.01	3.3	1, 136
(6) HELP	.58	.00	0.2	1, 135
(7) Two-way interactions[1]	.62	.04	0.4	25, 110

[1]All 25 product terms representing all possible two-way interactions among variables listed in (1) - (6) were entered as a set.

***$p < .001$.

Table 9

Results of Hierarchical Regression Analyses

for Computer-Administered Mathematics Test

Exlanatory Variables Added	Cumulative R^2	Increase in R^2	F for Increase in R^2	df
(1) Paper-and-pencil mathematics	.58	.58	199.8***	1, 143
(2) Paper-and-pencil reading	.62	.04	15.7***	1, 142
(3) Computer anxiety, computer confidence, computer liking	.65	.03	3.9*	3, 139
(4) Computer experience	.66	.00	0.9	2, 137
(5) Test order (practice)	.66	.00	0.0	1, 136
(6) HELP	.66	.01	2.6	1, 135
(7) Two-way interactions[1]	.71	.05	0.7	25, 110

[1]All 25 product terms representing all possible two-way interactions among variables listed in (1) - (6) were entered as a set.

*$p < .05$. ***$p < .001$.

Table 10

Regressions of Computer-Administered

Test Scores on Explanatory Variables

	Dependent Variable					
	Computer reading			Computer mathematics		
Explanatory Variables	b	SE	sr^2	b	SE	sr^2
Constant	1.58	3.25		3.88	2.75	
Paper-and-pencil:						
Reading	1.31	.16	.202	.54	.14	.039
Mathematics	.65	.17	.043	1.41	.15	.223
Computer Attitudes:						
Anxiety	.78	1.41	.001	-1.39	1.20	.003
Confidence	1.72	1.76	.003	4.22	1.49	.020
Liking	-.06	1.01	.000	-.52	.86	.001
Computer Experience:						
Exp. 1[a]	1.84	1.21	.007	1.13	1.03	.000
Exp. 2[b]	.10	1.43	.000	1.02	1.22	.002
Test Order[c] (practice)	-1.66	.91	.001	-.04	.77	.000
HELP[d]	-.14	.93	.003	-1.28	.78	.007
Multiple R		.76			.82	

Note. sr^2 is the squared semipartial correlation coefficient, which represents the unique contribution of the variable to prediction.

[a]Coded 0 if no previous computer experience and 1 if rare experience.

[b]Coded 0 if no previous computer experience and 1 if regular or routine experience.

[c]Coded 0 if computer reading test taken first and 1 if computer mathematics test taken first.

[d]Coded 0 if supervisor help available and 1 if no help available.

Table 11

Preference for Computer-Based and

Paper-and-Pencil Tests before and after Testing

Pretest Preference	Posttest Preference (%)			
	Computer test	Paper/pencil test	No preference	**Total**
Computer test (N=47)	85	6	9	**32**
Paper/pencil test (N=42)	36	38	26	**29**
No preference (N=56)	32	32	36	**39**
TOTAL (N=145)	**50**	**26**	**24**	**100**

Note. Question was "If there were a computer test and a paper-and-pencil test with the same questions, which would you prefer to take?"

Appendix A
Pilot and Study Samples' Responses to Computer Attitude Scale Items

Statement	Sample:	Agreed or Strongly Agreed (%)		Strongly Agreed (%)	
		Study	Pilot	Study	Pilot
Anxiety					
Computers do not scare me at all		65	76	23	46
Working with a computer would make me very nervous		30	22	6	3
I do not feel threatened when others talk about computers		66	70	28	43
I feel aggressive and hostile toward computers		16	9	3	1
It wouldn't bother me at all to take computer courses		71	79	30	49
Computers make me uncomfortable		42	24	6	3
I would feel at ease in a computer class		63	80	16	41
I get a sinking feeling when I think of using a computer		19	15	4	1
I would feel comfortable working with a computer		81	88	21	54
Computers make me feel uneasy and confused		37	19	5	2
Confidence Scale					
I'm no good with computers		47	22	12	4
Generally I feel OK about trying new problems on a computer		78	86	26	46
I don't think I would do advanced computer work		76	60	36	27
I am sure I could do work with computers		90	96	52	68
I'm not the type to do well with computers		39	20	6	4
I am sure I could learn a computer language		82	90	38	57
I think using a computer would be very hard for me		23	16	5	3
I could get good grades in computer courses		79	86	21	45
I do not think I could handle a computer course		13	9	7	3
I have a lot of self-confidence when it comes to working with computers		39	64	5	22
Liking Scale					
I would like working with computers		86	88	39	54
The challenge of solving problems with computers does not appeal to me		42	26	15	7
I think working with computers would be enjoyable and stimulating		72	85	29	44
Figuring out computer problems does not appeal to me		53	36	17	9
When there is a problem with a computer run that I can't immediately solve, I would stick with it until I have the answer		59	74	16	31
I don't understand how some people can spend so much time working with computers and seem to enjoy it		46	26	14	4
Once I start to work with a computer, I would find it hard to stop		38	57	8	20
I will do as little work with computers as possible		26	13	8	4
If a problem is left unsolved in a computer class, I would continue to think about it afterward		61	74	21	33
...ot enjoy talking to others about computers		47	33	10	7

Appendix B

Results of Regression Analyses for Two Subsamples who Self-Reported either SAT or ACT Scores

Table B.1

Correlations among Test Scores, Computer Attitude Scales,

and Computer Experience for a Sample of SAT Test Takers (N=50)

Variable	Mean	S.D.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Computer reading	27.0	8.1								
(2) Computer mathematics	26.7	9.1	.49							
(3) Computer writing	26.4	6.5	.52	.39						
(4) SAT verbal	513	95	.49	.31	.35					
(5) SAT mathematics	519	95	.25	.57	.15	.39				
(6) Computer anxiety	3.3	0.6	.23	.25	-.02	-.04	.25			
(7) Computer confidence	3.2	0.5	.13	.21	.01	.01	.36	.86		
(8) Computer liking	2.8	0.6	.07	.11	-.08	-.12	.14	.72	.81	
(9) Computer experience	0.2, 0.7	0.4, 0.4	.08	.21[a]	.16[a]	.15[b]	.17[b]	.51	.42	.42

Note. With a sample of 50, correlations of .279 are significant at the .05 level. Correlations of .361 are significant at the .01 level, two-tailed test.

[a]Both experience variables were weighted negatively.

[b]One experience variable was weighted negatively.

Table B.2

Correlations among Test Scores, Computer Attitude Scales,

and Computer Experience for a Sample of ACT Test Takers (N=141)

Variable	Mean	S.D.	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Computer reading	27.0	7.1							
(2) Computer mathematics	28.7	7.2	.66						
(3) Computer writing	27.1	4.7	.59	.48					
(4) ACT composite score	22.9	4.2	.55	.52	.47				
(5) Computer anxiety	3.3	0.6	.13	.20	.09	.24			
(6) Computer confidence	3.2	0.5	.18	.28	.13	.35	.84		
(7) Computer liking	3.0	0.5	.14	.21	.04	.25	.67	.72	
(8) Computer experience	0.3, 0.7	0.5, 0.5	.06[a]	.07[a]	.06	.15	.54	.47	.34

<u>Note</u>. With a sample of 141, correlations of .166 are significant at the .05 level. Correlations of .217 are significant at the .01 level, two-tailed test.

[a]Both experience variables were weighted negatively.

Table B.3

Results of Hierarchical Regression Analysis for

Computer-Administered Reading Test in Praxis Pilot Sample

Explanatory Variables Added	Cumulative R^2	Increase in R^2	F for increase in R^2	df
		ACT Test Takers (N=141)		
(1) ACT composite score	.30	.30	60.4***	1, 139
(2) Computer anxiety, computer confidence, computer liking	.30	.00	0.1	3, 136
(3) Computer experience	.31	.00	0.4	2, 134
		SAT Test Takers (N=50)		
(1) SAT scores (verbal and mathematics)	.25	.25	7.8**	2, 47
(2) Computer anxiety, computer confidence, computer liking	.34	.10	2.1	3, 44
(3) Computer experience	.39	.04	1.5	2, 42

** $p<.01$. *** $p<.001$.

Table B.4

Results of Hierarchical Regression Analysis for

Computer-Administered Mathematics Test in Praxis Pilot Sample

Explanatory Variables Added	Cumulative R^2	Increase in R^2	F for increase in R^2	df
		ACT Test Takers (N=141)		
(1) ACT composite score	.27	.27	52.0***	1, 139
(2) Computer anxiety, computer confidence, computer liking	.28	.01	0.7	3, 136
(3) Computer experience	.31	.02	2.2	2, 134
		SAT Test Takers (N=50)		
(1) SAT scores (verbal and mathematics)	.34	.34	12.0***	2, 47
(2) Computer anxiety, computer confidence, computer liking	.39	.05	1.2	3, 44
(3) Computer experience	.45	.06	2.2	2, 42

*** $p < .001$.

Table B.5

Results of Hierarchical Regression Analysis for

Computer-Administered Writing Test in Praxis Pilot Sample

Explanatory Variables Added	Cumulative R^2	Increase in R^2	F for increase in R^2	df
		ACT Test Takers (N=141)		
(1) ACT composite score	.23	.23	40.4***	1, 139
(2) Computer anxiety, computer confidence, computer liking	.23	.01	0.4	3, 136
(3) Computer experience	.24	.01	0.8	2, 134
		SAT Test Takers (N=50)		
(1) SAT scores (verbal and mathematics)	.12	.12	3.2*	2, 47
(2) Computer anxiety, computer confidence, computer liking	.13	.01	0.1	3, 44
(3) Computer experience	.14	.01	0.3	2, 42

* $p<.05$. *** $p<.001$.

53905-08877 • Y13M3.5 • 271823