Computer-Administered Testing for TASP®

John D. Mattar

The Texas Higher Education Coordinating Board (THECB), the Texas Education Agency (TEA), the Tests and Measurements Committee, and colleges and universities throughout Texas have been strong advocates of computerized testing for the Texas Academic Skills Program (TASP®). National Evaluation Systems, Inc. (NES®), has worked with these groups to design and implement a computerized test appropriate for TASP.

A primary goal of these groups has been to ensure that the delivery system for the computerized test is as fair and equitable to examinees as possible and is as compatible as possible with key program features. These goals have been achieved by developing and implementing a user interface that minimizes the impact of a computer administration on examinees, by producing a computer test that is comparable in content to the paper test, and by assuring the comparability of scores from the paper and the computer modes of testing.

Before I describe the actual test and its development, I would like to briefly address general concerns and issues related to the use of computers in assessment.

Recent research (Vale, 1995) proposes that there are two major reasons agencies pursue computerized testing: the convenient scheduling available to examinees and the immediate availability of test results. These advantages rely on the availability of convenient test locations and the economics of the test in comparison to the more commonly used paper-and-pencil versions of the test. In the TASP program, the large volume of examinees and the small number of different tests administered in the testing program make the computerized version of TASP feasible. Thus, the Computer-Administered TASP Test indeed serves the functions that are most important according to recent research.
The concerns Vale (1995) raised about computerized testing were two-fold. First, the results of the computerized version of the test must be comparable to the results of their paper-and-pencil companions. To determine whether the Computer-Administered TASP Test and the regular TASP yielded comparable results, NES conducted a pilot test. Computerization of a test should have as little impact on examinees as possible. Some earlier studies (Raffeld, et al., 1990) have shown a significant difference in examinee performance on computer tests compared to paper-and-pencil tests, with lower computer test scores. Our research into this issue for TASP yielded very different, and much more positive, results.

Second, the computerized version of the test should not create undue anxiety that may affect examinee performance on the test. The Computer-Administered TASP Test was specifically designed to be user friendly, even for examinees who are not comfortable with computers. Touch-screen technology—a large screen coupled with the absence of a keyboard and mouse—facilitates the use of this test. Since the test requires no computer skills, the Computer-Administered TASP Test is as easy to use as the paper-and-pencil version of the test. A survey, which I will describe later, was conducted to verify this assumption.

By describing the development and implementation of the Computer-Administered TASP Test and its hardware and software, I will illustrate how Texas educators and NES have addressed these concerns and issues and have contributed to advancements in the field of computer testing.

**Goals of Computer-Administered Testing**

Computer technology opens up new possibilities for test administration, particularly in terms of examinee access to testing and the time needed to score the test and produce the results. Computer-administered testing can serve as a useful complement to the regularly scheduled large-scale test administrations currently offered.
Daily access to testing. The administration of tests by computer can enhance access to testing by making the test available daily on college campuses. With tests stored on highly secure computers, access to testing can be made available on an ongoing basis. Expanded access to testing can be especially important for examinees who find themselves in emergency situations in which they must take the test at a time when a regularly scheduled test administration is not available. The availability of ongoing computer testing will increase the flexibility of the testing schedule to better meet the particular needs of students.

Score reporting timeframe. Computer-administered testing can reduce the time needed to score tests and produce score reports in two ways. First, since computer testing is ongoing, the number of examinees being scored at any one time is much smaller than it is for the regularly scheduled, large-scale test administrations. Second, since the examinee enters his or her responses directly into a computer, the multiple-choice portions of the test can be scored much faster. There is no need to collect, transport, and scan answer sheets. The writing samples still need to be scored by Texas educators and will follow a different timeline, which I will comment on later.

While taking advantage of what computer testing offers in terms of access and scoring, it was important, even critical, to keep three goals foremost in mind: to match the TASP Test design, to maintain the involvement of Texas educators, and to minimize the impact on examinees.

Matching the TASP Test design. There are many TASP program and test design issues that we needed to address in the development and implementation of a computerized test. The computerized test must incorporate key aspects of the test design, scoring, and score reporting found in the existing program. For instance, the hardware and software in use should make it possible to duplicate the content of the paper version of the TASP Test. This includes coming as close as possible to matching the format and "look" of the paper test for examinees. We also wanted examinees to be able to move forward and backward in the computer-administered test and to mark their answers as freely and easily as they would on the paper version. Finally, it was important to ensure that test results for both the
paper and computer versions are reported in the same fashion, regardless of when examinees take the tests or receive their scores. The only noticeable difference should be that the examinees who take the test on computer receive their scores in a shorter timeframe.

**Texas educator involvement.** As mentioned earlier, the THECB, the TEA, the Tests and Measurements Committee, and NES have worked together to design a system that meets the needs of Texas educators and Texas students and that is appropriate for the TASP Test. At each stage in the development of the computer-administered test system, these groups have discussed basic design issues as they relate to the needs of the state of Texas, the colleges and universities, and the students. The Tests and Measurements Committee reviewed the design of the examinee interface as well as the design of a pilot test study to determine the comparability of computerized test scores with paper-and-pencil test scores. The Committee has also been involved in discussions concerning the distribution of computer test sites, basic program policies, and registration and testing procedures.

Computer-administered testing was discussed with registrars, faculty, administrators, testing personnel, counselors, and advisors at a number of colleges and universities throughout Texas. Their concerns in such areas as distribution of computer testing sites across the state, access to computer testing for examinees from all colleges, fairness to examinees, score reports and score reporting timeframes, and space requirements at institutions have been considered in the implementation of the test.

**Minimize impact on examinees.** Minimal impact was the number one goal. Educators wanted the test on computer but also wanted virtually anyone to be able to take it. The TASP assesses an individual's skills in reading, math, and writing. No one wanted to add computer skills to the list of skills an examinee needs in order to pass the test. The TASP population is diverse in terms of its age and its experience with computers and access to them. Everyone wanted to devise a testing system that makes the computer aspect of the test of no adverse consequence to examinees.
TASP Test Parameters and Test Validity

In order to be a valid test, the computer-administered version of the TASP Test must match the test design developed and approved by the Content Advisory Committees, the Bias Review Panel, the Texas Higher Education Coordinating Board, and the Texas Education Agency. The TASP Test includes many special features that must be incorporated in the computer-administered delivery system. The TASP Test given on computer must be comparable to the existing TASP Test since the results from both modes of administration will be used interchangeably. Any change to the test design parameters would represent a potential threat to the validity of the TASP Test.

Test form construction and the skill-level test blueprint. One key component of validity is the skill-level blueprint that is used to construct each test section. The TASP Test is a criterion-referenced test designed to assess particular skills. The TASP skills were developed and validated by thousands of Texas educators and are the basis of the entire program. The skills define what students need to know and be able to do in order to perform effectively in college-level work. The distribution of items that appear on a test across the TASP skills is a critical feature of the test design and of the validity of the test. Each form of the test must meet the skill-level distribution requirements so that each form has appropriate content representation and is comparable in this regard to other forms. If this blueprint is not met, the test form is not a valid measure of the TASP skills.

Diagnostic reporting. An important outcome of the skill-level blueprint of the TASP Test is the ability to report diagnostic feedback to students, faculty, counselors, and advisors. Diagnostic reporting at the skill and skill-area level can be accomplished only if the test forms represent the TASP skills in the appropriate proportions. Such diagnostic reporting is a key feature of the TASP program since the program is designed to support access to education and to appropriate remediation for students.

Reading Test passages. The TASP Test is designed to assess examinee reading skills in relation to the types of reading tasks that college students face each day as part of their coursework. Accordingly, the TASP reading test uses extended reading
passages of up to approximately 750 words in order to represent the types of reading that college students must face in college textbooks, journals, and other college course materials. Each reading passage is accompanied by six multiple-choice items, each item measuring one of the six TASP reading skills. The TASP Content Advisory Committee felt very strongly that the reading skills should be tested in concert with each other in a realistic reading passage. The committee felt that reading skills should not be tested in isolation in brief reading selections. For each passage, the six items are presented in a fixed order by skill number, as determined by the test development committees. The use of extended passages with each passage accompanied by items representing all six reading skills is a key design feature of the reading test, as defined and approved by Texas educators.

**Mathematics Test.** The mathematics test includes extensive, detailed graphic elements (graphs, charts, tables, diagrams, symbols) as well as four pages of definitions and formulas for use by examinees. These elements are included to represent the range and complexity of mathematics skills and problems that college students face. Graphic elements are a key feature of many of the items measuring the validated TASP mathematics skills. In some items, each of the four response options is a complete graph or chart, so a single item may contain up to four complete graphs or charts. When the mathematics test was updated and revised in 1993, the TASP Content Advisory Committee felt very strongly about the addition of more skills and items involving graphic elements.

**Writing Test passages and writing sample section.** The writing test includes passages of college-level text of up to approximately 200 words. These passages help create a realistic context for asking questions about effective writing. Each passage is accompanied by two or three multiple-choice items. The TASP Content Advisory Committee felt strongly that questions about writing skills should be asked in relation to an authentic reading passage that contained a number of errors as well as many elements without errors. The committee was clear that it did not want to test for error recognition in isolated sentences.
The writing test also includes the writing sample section, containing a prompt to which the examinee must respond in writing. Both the passages (with multiple items) and the writing sample section are key design features of the test, as defined and approved by Texas educators.

These reading, mathematics, and writing features of the TASP Test, along with many other key design features, were successfully incorporated in the computerized version of the test.

Compatibility in score reporting to institutions and examinees. In designing a system for reporting the TASP scores, we recognized the importance of maintaining the existing score reporting structure—of ensuring that use of multiple modes of testing would not add a level of complexity to the score reporting process. This is especially critical for the TASP Test score reports because they include detailed skill-level diagnostics. Maintaining a single structure for score reporting helps to ensure that students, advisors, and counselors get consistent information about student performance, whether the test was taken on computer or paper and whether the scores are reported in 15 minutes, in one week, or in three.

Computer-Administered Test Delivery System Features—Hardware and Software

As noted previously, in a high-stakes testing situation any computer technology in use should be as user friendly as possible. The mode of test administration (e.g., computer delivery) should have as little impact on examinees as possible. The computer-administered version of the test should not introduce new requirements for passing the test (e.g., the examinee should not need to be familiar with computer operations in order to pass the test). The TASP testing population is very diverse in terms of age, computer experience, and access to computers.

Impact on examinees. If the delivery of a test on computer is too complicated for examinees, significant amounts of examinee practice and training are required before a test can be taken. This situation would clearly add to testing time, tie up valuable hardware and testing space, and also significantly increase test
anxiety for examinees. If an examinee needs to complete a lengthy computer tutorial just before taking a test, that examinee may experience increased anxiety; and such a testing system, which favors examinees with prior computer experience, is potentially unfair to those examinees who have not had much (or any) prior computer experience. Many TASP examinees have little or no direct experience with computers, especially the many students who are returning to higher education after a long hiatus from formal education. We focused on that fact when selecting hardware and developing software for the computerized version of the test.

Hardware

**Large screens.** A key feature of the system is the use of 20-inch screens (as compared to the standard 14-inch screens found on most computer monitors). The surface area of a 20-inch diagonal screen is nearly twice as large as the surface area of a 14-inch diagonal screen.

The larger screen provides two key advantages. First, the larger screen accommodates more of the long reading passages and extensive mathematics graphics, thereby minimizing the need for examinees to scroll through material in order to answer test questions. Second, the larger screen allows for higher resolution graphic elements than can be presented on a standard 14-inch screen. Graphic element resolution is particularly critical in the mathematics test.

Higher resolution text and larger character sizes are also possible. Higher resolution text may be easier for many examinees to read, especially examinees with learning disabilities and older examinees. Ease of reading on the screen may be an issue for examinees, since they will need to read from the computer for an extended period of time. The availability of larger fonts also increases the options for accommodations for the visually impaired.

**Touch-screen technology.** Another significant feature of the examinee interface is the use of touch-screen technology. The examinee completes the test simply by touching designated areas on the screen. No computer skills are required. The
Examinee does not need to be familiar with computer operations and does not need to have keyboarding or mouse skills in order to take the test.

**Minimized need for training and examinee computer skills.** The combination of large screens and touch screens minimizes the need for prior computer skills and the need for extensive training of examinees at the test site. This is a key component in minimizing the impact of the computer test delivery system on examinee performance.

**Software**

**Graphical user interface (GUI).** The test delivery system is based on a graphical user interface that simplifies the task of taking the test on computer. Commands are input by simply touching a part of the screen. No specialized, separate inputs or commands are needed. There are no "hidden" commands. Directions and inputs appear on the screen in a user-friendly format in plain language. For example, each touch area on the screen is clearly labeled with self-explanatory instructions or labels such as "next item," "previous item," "score test," etc. For example, to scroll down a passage, the examinee touches an arrow that points down; to scroll back up, the examinee touches an arrow that points up; to choose a response, the examinee touches that response.

**Examinee help.** The "help" function is context sensitive: to get "help" information at any time during the test, the examinee need only touch the designated area on the screen, and help information relevant to where the examinee is at that time will appear. In addition, a test administrator is available throughout the test period to answer examinee questions, such as how to use the commands, indicate their answers, or move around different parts of the test.

**Minimized need for computer training for the test.** As with the hardware configuration, the software features of graphical user interface and the examinee "help" function minimize the need to train examinees on the computer before the test begins.
Reviewing answers, previewing questions, skipping questions. The system allows examinees to review their work and change their answers as they see fit. The examinee can change his or her answers to questions at any point until the examinee indicates that he or she is finished or the testing period officially ends, whichever comes first. The system also allows examinees to preview one or all of the test questions before answering them. Examinees are also able to skip over a question, with the choice of returning to it later, if desired. These capabilities are all features that examinees are accustomed to with paper-and-pencil tests.

Answer sheet. In many software applications there is a clear relationship between non-computer and computer applications. For example, in some personal organizer applications, the software displays a simulated datebook, calendar, and telephone/address list in everyday formats that are familiar to users. As an extension of this idea, the "answer sheet" feature built into the Computer-Administered TASP Test allows an examinee to review his or her answers at a glance and to return to any specific question simply by touching the number for that question on the "answer sheet." This "answer sheet" looks like the answer sheet that examinees are familiar with from any standardized test.

To answer a question during the test, examinees simply touch one of the response options offered in that item. That response option is then darkened by the computer, indicating that it was selected as the correct response. At the same time, that answer choice is automatically transmitted to and recorded on the "answer sheet." Examinees select and change their answers only on the screens displaying the actual test questions. It is important to emphasize that an examinee cannot change an answer directly on the "answer sheet." The examinee must return to the question in order to change the answer. This feature guards against accidental answer changes. Examinees can, however, view their "answer sheet" at any time. Viewing it is, among other things, a quick way for examinees to identify and go to any question they have intentionally, or unintentionally, not yet answered.

Instant feedback. When an examinee finishes the test, a preliminary score report for the reading and mathematics sections will be printed for the examinee. This report will contain the
scaled score for each section, the skill area performance indicators (diagnostic reporting), and a digital photograph of the examinee. In this way, examinees will receive immediate feedback concerning their performance on the test. For the writing test, the main determinant of a score and passing status is the writing sample. A score for the writing section will be available one week after the essay is received by NES in Austin. It would be possible to report the skill area performance for the writing multiple-choice questions on the day of the test, but a scaled score and passing status cannot be reported because the writing sample is the key factor in passing status.

Pilot Testing

The THECB, the TEA, the Tests and Measurements Committee, and NES have worked together to design a pilot study of the Computer-Administered TASP Test.

Ensuring comparability of computer scores with paper-and-pencil scores. As discussed above, a key issue in computer-administered testing is to ensure that scores from a computer-administered version of a test are comparable to scores from a paper-and-pencil version of the same test. This is crucial since both versions of the test will be available to students. In order to ensure fairness for all examinees, results from the two versions must be comparable, just as two different forms of the paper-and-pencil test must be comparable. Scores from the two versions will be used interchangeably. A college or university should not have to be concerned with whether an individual student took the test on computer or on paper. Examinees who take either version of the test should be on equal footing; one group should not be unfairly advantaged or disadvantaged in relation to the other group.

As noted previously, some earlier studies of this issue with other computer systems and other tests found that the use of a computer had a negative impact on examinee performance (i.e., examinee scores on a computer version of a test were significantly lower than examinee scores on a paper-and-pencil version of the test). Some researchers attributed these differences to the need for computer skills and experience with computers. Students without much prior computer experience
may have spent more time dealing with the unfamiliar computer than answering the questions. Lengthy and involved training sessions just before a test may not only be difficult for an examinee to absorb but may also create undue test anxiety in examinees who are unfamiliar with computers and computer operations.

**Pilot test sites.** The pilot test was conducted at three institutions in Texas—Austin Community College, Southwest Texas State University, and University of Texas at San Antonio.

**Pilot test design.** The pilot test was conducted using two forms of the PreTASP Test (PTT®). The PTT is designed to mirror the TASP Test closely in terms of test construction and level of difficulty. Test questions for the PTT have been part of the regular TASP Test development process and have been pilot tested on operational TASP Test forms.

One hundred sixty-nine students completed the pilot test. Each examinee took two versions of the PTT, one on computer and one on paper. Some students took the test on computer first, others took the test on paper first. Students did not take the two versions on the same day. Each student was assigned randomly to one of four test conditions.

- **Condition 1:** First form 03 computer, second form 04 paper
- **Condition 2:** First form 03 paper, second form 04 computer
- **Condition 3:** First form 04 computer, second form 03 paper
- **Condition 4:** First form 04 paper, second form 03 computer

**Pilot test data analysis.** Examinee scores were studied using an analysis of variance procedure (ANOVA). A two-by-two ANOVA was used to analyze for effects of test method (paper and computer). In addition, demographic information and background characteristics (experience with computers, etc.) of examinees were collected and summarized.

**Examinee interface with the computer delivery system.** Examinee feedback on the computer delivery system was collected as part of the pilot test. Examinees responded to both selected-response and open-ended questions about the computer delivery system. These questions were designed to elicit feedback about important aspects of the system and to gather information about examinee experiences and prefer-
ences. Students were asked about the monitor screen, glare on the screen, text contrast, the touch-screen interface, the method of marking answers, the experience of reading questions on the computer, testing time, etc.

The computerized test system must be appropriate for the testing population. It must not have a negative impact on examinee performance or perceptions. Enhancements to the system (both hardware and software) were made based on examinee feedback.

**Pilot test data summary.** The analysis of variance procedure yielded no significant differences between examinee scores on the computer-administered version of the test and the paper-and-pencil version. Mean reading and mathematics scores were compared in the analysis, and the results support the interchangeable use of results, whether examinees take the test via computer or in a traditional test setting.

**Implementation**

With the pilot test completed, additional computer testing sites were established across the state. A registration system was developed, allowing candidates to register by phone up until the day before testing, on a space-available basis. Since January 1996, hundreds of students have taken the TASP Test on computer. For more information on this mode of TASP testing, see the Appendix for a copy of the TASP Registration Bulletin Supplement for computer-administered testing.

**References**
