

# Electronic Versus Paper-Based Testing in Education

Stephanie D. Holder  
American University  
Washington, DC 20016 USA

and

Rick Gibson, Ph.D.  
Department of Computer Science and Information Systems  
American University  
Washington, DC 20016 USA  
rgibson@american.edu

## Abstract

The introduction of computers into classrooms has provided most educators with the ability to use computer-aided, electronic tests for their students. However, there are issues and concerns related to computer-aided tests, which have a different look and feel (interface) in comparison to the standard paper and pencil format used in past years to test student knowledge. Instead of a number 2 pencil and a bubble sheet, students today are often presented with a monitor and a mouse in the computer-based testing (CBT) environment and asked to submit answers with a click of the mouse while reading questions on a computer monitor. Questions concerning validity and the reliability of computer-aided tests will be discussed as well as electronic testing trends.

Keywords: Benefits to Educators, Computer-Adapted Test

## 1. BACKGROUND

Are computer-aided tests a better choice over the paper and pencil format in testing student populations on their knowledge? In an attempt to answer this question, this paper will provide:

- Current information on the research that supports the benefits of using computer-aided testing and the use of these tests as a reliable and valid method (Saggio, 1996) of measuring student knowledge.
- A discussion of issues regarding computer-aided testing and the added beneficial value (Netlabs, 1998) it offers over paper and pencil testing.
- Benefits to educators is information on the accuracy in answer gathering, analysis of scoring results and the reduction of human error in test taking, as well as computation and analysis (Meijer, 1999) of the test results on the computer as compared to paper and pencil testing formats.

- An examination of work conducted in the analysis of computer-aided testing (Mills, 1999), but will be limited to the results of the tests and not the analytical methodology of the tests.
- Data that reflect the benefits, trends, and potential pitfalls of converting from paper and pencil to computer-aided testing.

Russell and Haney (1997) observed that two of the most prominent movements in education over the last decade or so are the introduction of computers into schools and the increasing use of authentic assessments. A key assumption of the authentic assessment movement is that instead of simply relying on multiple choice tests, assessments should be based on the responses students generate for open-ended real world tasks. At the state level, the most commonly employed kind of non-multiple-choice test has been the writing test (Barton & Coley, 1994) in which students write their answers longhand. At the same time, many test developers have explored the use of computer administered tests, but this form of testing has been limited almost exclusively to

multiple-choice tests. Relatively little attention has been paid to the use of computers to administer tests which require students to generate responses to open-ended items.

Russell and Haney (1997) consider it likely that increasing numbers of students are growing accustomed to writing on computers. Nevertheless, large-scale assessments of writing are attempting to estimate students' writing skills by having them use paper-and-pencil. Their results, if generalizable, suggest that for students accustomed to writing on computer for only a year or two, such estimates of student writing abilities based on responses written by hand may be substantial underestimates of their abilities to write when using a computer. This suggests that educators should exercise considerable caution in making inferences about student abilities based on paper-and-pencil/handwritten tests as students gain more familiarity with writing via computer. And more generally it suggests an important lesson about test validity. Validity of assessment needs to be considered not simply with respect to the content of instruction, but also with respect to the medium of instruction. As more and more students in schools and colleges do their work with spreadsheets and word processors, the traditional paper-and-pencil modes of assessment may fail to measure what they have learned.

Although it will be some years before schools generally, much less large scale state, national or international assessment programs, develop the capacity to administer wide-ranging assessments via computer. In the meantime, we should be extremely cautious about drawing inferences about student abilities when the media of assessment do not parallel those of instruction and learning.

Before converting paper and pencil tests to computer-aided or computer-adapted formats concise research is needed to provide the knowledge available on the success and benefits of using computer-aided tests as well as potential pitfalls (McBride, 1998). Traditionally most test administrators have formulated test questions and formatted those test questions on paper to be presented to students for testing purposes. Students taking tests in paper and pencil format prior to the advent of computers were administered tests in paper and pencil format.

The standardized test is often created by commercial test publishers and is crafted to enable educators to give large numbers of students throughout the country. These standardized tests give educators a common yardstick (Bagin, Rudner 1993) or "standard" of measure. Generally students were required to select the correct answer by placing an 'X' in a box, shading in a circle with a #2 pencil, or writing the answer in a blank space on the test sheet. The practice of using paper and pencil testing has been the standard form of assessing student knowledge of subject matter for many years

(Netlabs, 2000), and is still in use today. Many educators within the United States rely on paper and pencil format especially when administering standardized testing, which dictates that the testing content and environment are identical for all test subjects. Paper and pencil testing has aided educators by allowing them a means of equating the student's level of knowledge in a way that is simplified and easy to fairly administer, although at times time consuming. Students are gathered in a classroom setting at a specified time and are all tested at once using the same format and list of questions. Standardized tests can be used to assess student knowledge across many subjects some examples include law, medical science, and business. Tests for these subjects include the LSAT, MCAT and GMAT (Saggio, 1996). On the high school level the SAT and ACT are examples of standardized tests used to assess student knowledge in preparation to enter college (Bagin, Rudner 1993). In addition many licensing and professional exams are given in paper and pencil format.

For example, The Commission on Dietetic Registration (CDR) is the credentialing agency for The American Dietetic Association. The purpose of the Commission is to protect the nutritional health and welfare of the public by establishing and enforcing certification and recertification standards for the dietetics profession. The Commission made the decision to implement computerized testing for the entry-level examinations because it recognized the many advantages it offers to examinees. These include: Flexible test administration dates; examinees can schedule testing throughout the year, rather than the current two test dates per year; Retesting available six weeks following the previous test date; Unique examination based on each examinee's entry-level competence; Score reports distributed to examinees as they leave the test site eliminating the six week waiting period required with paper and pencil testing.  
(<http://www.cdrnet.org/certifications/rddtr/cbtfqa.htm>)

Educators also use standardized tests as a measure as a methodology for human subject psychological assessments testing and skills based testing. Weisband and Keisler (1996) conducted a meta-analysis that gives support to the main hypothesis that computer administration elicits more self-disclosure than traditional forms of administration do. In recent years, the disclosure effect has declined significantly due, perhaps, to increasing public knowledge of computers, increasing public computer literacy, or even people's reduced awe of the computer. Their indirect analysis, however (for example, comparing students to other adults), did not support explanations related to public knowledge of computers.

An unresolved research issue is that changes in computer technology have made it possible for a

computer instrument to have the "look and feel" of a paper-and-pencil questionnaire, typed form, or printed test. Forms now look more like paper questionnaires, forms, or tests than they did in earlier years, and allow for more stereotypical questionnaire-type responses using radio buttons and fill in blanks (as compared with typed commands). These changes might have increased respondents' sense of the computer interaction as an evaluation or test situation and consequently reduced their disclosure. Unfortunately, we were unable to evaluate this idea in the meta-analysis because few investigators described their computer interfaces in sufficient detail. Possibly researchers did not realize computer interfaces would change so much. In any case, the idea that differences in the interface can affect disclosure has not been investigated yet. A related issue is the belief by many that answering questions on a computer changes respondents' perceptions of the test environment. For example, working on a computer could create a sense of privacy or anonymity. Some investigators have reported a strong relationship between anonymous and identified computer response.

However, research is needed for the following issues:

- Ethical Issues-- If people have a false illusion of privacy or otherwise let down their guard when they respond to a computer, the world has discovered an easy, cheap way to obtain sensitive information from people. Currently, the American Psychological Association Guidelines for Research on Human Participants, as well as most research organizations' codes of research conduct are silent with regard to such topics as how to obtain informed consent electronically (and whether it is legitimate to do so), how much to reveal about remote sites of data collection, and about electronic forms in general.
- Policy Issues--the use of computer forms has not proved to be a source of social disagreement in the way computerized monitoring and informal electronic communications like email have. The legal situation is presently in a state of flux. To the degree that people believe or perceive their communications through computers to be safe, current organizational and legal policies may be inappropriate.
- Design Issues--As the power and speed of computers continues to increase, researchers and technologists have responded by improving the readability and ease of response in computer forms, interviews, surveys, and tests. Advances in computer interfaces also have increased the variety, credibility, and salience of social information in forms, for example, through animated characters or icons. Will new computer interface designs induce people to act more like they would using a paper questionnaire or in an interview with another

person? Computer forms are multidimensional, increasingly so as interfaces incorporate speech and speech recognition, auditory and kinesthetic feedback, social intelligence, emotional response, directed animation, talking to people on the screen or even virtual reality. To understand how design affects disclosure in new computer instruments, we will have to investigate which features of computer forms affect people's perceptions and responses. It will be interesting to see how much users disclose when the computer form is delivered by an animated cartoon character.

With the advent of computers entering the testing arena standardized tests have become computerized and are of benefit to educators as an efficient method of testing in terms of the amount of time spent on testing students. Large groups of students can be tested at one time (Bagin, Rudner 1993) in different cities or classrooms, and a standard of measure can be obtained in a short amount of time. Due to the strength of standardized tests they will most likely remain the choice of educators to assess student knowledge and the new computer-aided formats will ease the time and effort needed to test students. The speed with which computers can quickly assess student knowledge, formulate test questions, and score tests through computer algorithms is unparalleled by human effort. Computer programs can generate questions randomly, assess the student's knowledge by adapting questions to their level of subject accomplishment, and score the exams at the same time the student is taking the test. Due to the significant benefits of computer-aided tests educators and institutions are leading toward the development of computer-aided standardized tests to replace their paper and pencil counterparts.

The ETS testing organization for the GRE, GMAT, LSAT, and MCAT is gradually phasing out the paper and pencil versions of its tests in favor of the computer-aided versions (Saggio, 1996). Professional licensing exams such as the U.S. Medical Licensing Examination ([Acep.org.htm](http://Acep.org.htm), 1999) is also adapting the use of computer-aided test over paper and pencil due to the acknowledged benefits of the new format. In the medical licensing and science fields the benefits of computer-aided tests can have additional assets for the educator in not only are the tests as reliable and valid in testing student knowledge they also provide built in safety measures such as computer simulated patients ([acep.org.htm](http://acep.org.htm), 1999).

## 2. DISCUSSION

In reviewing the current research in the field of standardized testing it is apparent that the standardized test may be one of the more common types of tests used in student populations. In comparison to the oral or essay test, the standardized test can be faster to administer (Arlin, 1982). The test is also easy to give to

large groups of students. Standardized tests are generally fixed-length computerized (or paper and pencil) exams which present the same number of questions to each test taker. The score of this type of test is usually depends on the number of questions answered correctly in the case of the GRE and GMAT (Ji, 1998). The SAT test is designed to derive how well the student is doing in comparison to other students taking the same test (Bagin, Rudner 1993).

Traditionally standardized exams have a successful history dating back to the second decade of the 20<sup>th</sup> century (Dpo.htm, 1998). One of the challenges of the traditional paper and pencil standardized test relates to the amount of questions a student needs to be presented with in order for the educator to get a clear picture of how much knowledge the student has on the subject being tested. The traditional paper and pencil tests often present more questions than are necessary to assess a student's knowledge level. Answering easy questions correctly doesn't tell us much about the student; many students can answer easy questions. In the same manner difficult questions answered incorrectly also doesn't tell the educator much about the student's level of knowledge. A better testing solution would be a method available to discover the level of a student's knowledge on a scale that ranges from easy to hard questions. A score or 'standard' could be derived for that level (Bagin, Rudner 1993). The Computerized Adaptive Test (CAT) is a good choice. It is a test that quickly finds the level of knowledge of the student taking the test. In terms of time saving in the classroom or testing center the (CAT) is a better choice than the paper and pencil test. An added advantage in using (CAT) is that the student is less likely to get bored with a shorter test, and students on average are found to like computerized exams (Oglilivie, 1999).

The history of adaptive testing starts in the beginning of the 20<sup>th</sup> century when large-scale testing began (Dop.htm, 1998), one of the first adaptive tests constructed was created by Alfred Binet. Binet's IQ test began with a questions matched to a child's age, and ended when the child could not answer a few questions in a row with the correct answers. Although the original Binet IQ test was not computerized until later in the 20<sup>th</sup> century, the test is still today considered reliable and valid at judging a child's IQ based on age. Computer adapted tests work the same way paper and pencil adaptive tests work, the test adjusts to the ability of the test taker. The main difference in the paper and pencil adaptive test and the computerized version is speed and efficiency.

The (CAT) can compute the student's score with fewer questions, decreasing the testing time (Wise, 1989). The benefits of saving time administering standardized tests serves two important purposes, first it allows educators a way to judge the students ability to process questions and secondly it frees the test environment for another

group of students. Today many institutions for example Education Testing Services (ETS) considered (CAT) a reputable methodology for judging student knowledge and have adopted the (CAT) test format as a primary way to administer tests.

#### **Use of computer adapted test format in leading standardized tests**

In 1998 the Graduate Record Exam (GRE) began initial work (Saggio, 1996) on a computer adapted test (CAT) version of the GRE. The GRE is an exam that is used as a measure to determine the knowledge of a prospective candidate for the purposed of entry into graduate school. Traditionally the GRE was presented in paper and pencil form. The GRE is given nation wide and considered a valid, reliable way to test a student's skill level. As of this writing the GRE tests can be taken in paper and pencil or (CAT) version, however the trend of these tests is to convert them to (CAT) on a wider scale. In comparison to its computerized counterpart the paper and pencil GRE allows students to erase a filled in circle answer and change to another answer. In the (CAT) version of the GRE "since the computer scores each question before selecting the next one the test taker must answer each question when presented and move on, test takers may not go back and change answers (GRE, 1998). The format is the same for other (CAT) tests as well, the National Council Licensure Examination is given in (CAT) format, and allows test takers no control over the ability to change answers (Steadman, 1998). The use of computerized adaptive testing (CAT) has increased substantially (Meijer, 1999) since it was first formulated in the 1970s. Research on test taking behavior of students it has found that student populations show the ability to adapt to the demands of testing no matter what realm of the testing situation and so as well on (CAT) tests where they cannot review questions or change answers (Stedman, 1998).

#### **Current trend of using computer adapted testing**

Many universities require standardized tests and few of the more traditional tests are the Graduate Record Exam (GRE), Law Standardized Adaptive Test, (LSAT), Medical Computer Adaptive Test (MCAT), and the test used to test high school students the Standardized Adaptive Test (SAT). In many cases student must take these tests to be considered for admission to high schools and universities. Educational institutions rely upon the aforementioned standardized tests as a valid way of judging student skills knowledge on subject matter prior to admission to U.S. educational institutions. As of this writing, Educational Testing Services (ETS) administers the GRE, LSAT, and Sylvan Centers administers the SAT to thousands of students applying to colleges and universities throughout the United States. ETS and Sylvan's standardized paper and pencil tests are looked upon as accurate by educators in measuring knowledge skills and are considered by educational institutions to be valid in testing skills and reliable over time and subject matter. Some

organizations in the test administration business are doing well and expanding like Sylvan (Brin, 1999). Other testing institutions like ETS are starting to retire a few of its paper and pencil tests, the GRE paper test is due to retire in the U.S. by 1999 (GRE, 1999) and U.S. students will be required to take the (CAT) version. Albeit standardized testing formats for students will likely be around for some time in the future and will continue to be used widely by educators especially in (CAT) format.

#### **An overview of computer-aided test format**

It is important to distinguish computer adapted test format from that of computer-aided. Some Computer-aided tests for example Self-Adapted Tests (SAT) can give the test taker more flexibility is going over answers or returning to particular parts of the test questions or answers. In a research study conducted to evaluate the ability of students to adapt to computer-aided tests the researcher found in the study of (a) the measurement precision and efficiency, (b) the effects of several individual difference variables (test anxiety, verbal self concept, computer usage, and computer anxiety), college students indicated, that they prefer tests in which they have as much control over the test (Wenzel, 1999) and as much information as possible. Computer-aided tests like the Self-Adapted Test (SAT) allow the test taker more control (Vispoel, 1999). This can be an important factor for educators in choosing the computer-aided test over computer-adapted. Computer-aided tests can also be valuable to educators for skill based testing. Typing tests, computer proficiency, and even simulated motor vehicle driving tests can help educators assess the knowledge of college students in an efficient safe manner. Another benefit is that many student like computer-aided tests and feel less anxiety (Wise, 1999) when taking them over paper and pencil tests.

#### **Costs and benefits of using computerized testing formats**

There are compelling reasons to convert paper and pencil tests to computerized (CAT) or computer-aided tests format. First there is the accuracy of the results. Computer algorithms take care of scoring of computerized test is a fast accurate manner. Tests scored on computer especially for large numbers of students are much less prone to human error in scoring (dop.htm 1998). Because computers can score test in a fast manner, for many computer-adapted tests like the GRE the scoring is in real time as the student answers the questions with the click of a mouse. In the educational environment of a classroom, or computer lab located on the schools grounds the need for professors to spend time proctoring an exam is reduced as lab workers can oversee the computer room test environment. Another benefit is the availability to test large groups of student at one time in a fast manner allowing for turnover of the test environment. Popular tests like the GRE (CAT) test format support nice features like getting a printout of your scores in a fast

manner, no longer do test takers have to wait a week for scores (Saggio, 1996). Like testing centers, educators can use the schools computer facilities or computerized classrooms and allow TA's to proctor computer-aided tests. This can save time and money in the reduction of man-hours needed to proctor exams.

#### **Benefits to educators using computer-aided tests for distance learning students**

As institutions seek new and enhanced ways of testing students another arena for computer-aided examinations may lead educators to apply this format to on-line distance learning exams. On-line students can access exams by logging on to the World Wide Web. In many cases it may be mandatory that students take tests for their distance learning classes on-line. If on-line distance learning tests are to accurately reflect classroom paper and pencil exams knowledge will be needed by educators and test developers to convert paper and pencil test to computer-aided formats and avoid any pitfalls (McBride, 1998) in test conversion. It is important to the correct choice in selecting computerized test formats. The computer-aided test may be better over a computer-adapted version for some exams.

As computers have entered the arena of standardized testing, test formats have evolved over time, to better provide test administrators with new technologies for testing. As U.S. classrooms become computerized, access by students to computer-adapted tests has the potential to be of benefit to educators. The trend of converting to computer-aided testing by large testing companies like ETS suggests there are benefits in using this format over traditional paper and pencil tests. With regard to the cost of computerized testing, it is beyond the scope of the paper, however it is a consideration in the adoption of computerized testing as a factor. The benefits of computer-aided testing in the end should be justified. The research presented shows evidence that supports the acceptance and trend of using computer-aided testing over pencil and paper testing.

There are many advantages and benefits of using computer-aided tests over paper and pencil tests, some examples are: (a) immediate scoring and feedback; the test taker can immediately know the results of the test; (b) unbiased scoring, computers score everyone the same; (c) accurate scoring, much less potential for human error than pencil or paper; (d) increased efficiency, you can test with less questions saving time; (e) convenience, computerized tests can be at times and testing centers convenient to the test taker; (f) improved test security, random ordering of test questions, adaptive testing, and selective questions make it difficult for test takers to cheat and use the answers of others; (g) reduction in answering errors, test takers can miss filling in a bubble on a paper and pencil test, however a computerized test can warn the test taker to fill in the answer; (h) computerized test questions can be changed easily; (i) computerized test use algorithms for test

question selection routines; (j) computerized test on average are less expensive, (dop.htm, 1998) to administer, and develop.

The paper presents the reader a view of current information on the benefits of choosing compute-aided testing over paper and pencil test formats. The purpose of this review is to bring to the reader additional information on the value of computer-aided testing, and bring to the reader reasons for accepting the use of computer-aided tests. It is of importance for educators to understand the relevance of using computer-aided testing formats to enhance current methods of assessing student subject knowledge. Although the paper is limited to comparison studies and benefit appraisals of computer-aided tests, educators were made aware of selected current studies on the risks of computerized test as well. For information on the potential bias of computerized tests on minority populations (Asian, Hispanic, African, European-Americans) which is beyond the scope of this paper refer to (Ji, 1998).

Educators will be able to address some of the issues of computer-aided testing with detailed findings on what to expect in terms of the time and cost saving benefits of computer-aided tests. It is relevant to bring into focus the differences students' face when administered computer-adapted tests due to the trend of their use in standardized tests such as the GRE and GRE subject tests. The trend of computer-adapted tests may continue and increase as its format is intended to assess a student's knowledge level in a less time than the computer-aided format. Although the length of test taking time can reduce time in the classroom setting, it is of great importance to educators to research computer-adapted testing further to insure (CAT) test questions derived from paper and pencil format are valid after conversion. Due to the ease and simplification of converting test questions many educators may tend to rely on their own expertise in selecting paper and pencil test questions for (CAT).

It was not the intention of this paper to cover test conversion to (CAT) rather to bring the readers attention to the diversification of computerized testing and its potential benefits. It should be noted however that test conversion of paper and pencil tests to computer-aided versions could be crafted easily using computer programs or purchased by educators from test publishers. Care in selecting computerized testing formats is needed to achieve the desired benefits of the test. Factors in selecting computerized test formats include various forms such as computer-aided tests; computer-adapted tests, self-adapted tests, and computer based tests that are generally used for testing skill knowledge.

Earlier studies in computer-aided and computer-adapted tests have in general been conducted on standardized tests including the Graduate Record Exam (GRE) and

Law standardized adapted test (LSAT). There is empirical data supporting the benefits of using the (CAT) for GRE testing which include information on the success of the test format (Mills, 1999) and there for the move to use the computer-adapted GRE as the sole measure. The paper and pencil GRE general test was retired April 10, 1999 (GRE, 1999). ETS administers a paper and pencil version of specific subject tests these tests include: Economics, Science, and Psychology to mention a few. Reports on (CAT) show it is possible to assess student knowledge and as an added feature shorten the test taking procedure by presenting what are considered to be harder more complex questions once less complex questions with easier answers are obtained. For the educator (CAT) tests can result in testing student knowledge in less time that using paper and pencil or computer-aided tests. Computer-adapted testing can be time effective for educators. The time spent in making decisions on (CAT) test questions can be scaled down through the use of computer programs designed to select test questions. One of the benefits of scoring the (CAT) is the reduction in time as compared to scoring a paper and pencil test by hand. Educators can use the time saved on test administration on other educational tasks for students. There can be the benefit if more time being spent on teaching new material than taking a large chunk of time during the class period for paper and pencil exams.

The paper reflects the opinion of some current researchers that computer-aided testing is a benefit and just as effective as paper and pencil tests to assess student knowledge. The implication of this review is to bring to light the ever-increasing use of computer-aided testing across the field of education. Research also supports the opinion that computerized testing may be the wave of the future for educational institutions as reflected by ETS and Sylvan Learning Centers, both large testing organizations that are moving toward a trend of using computer-aided and computer-adapted testing in place of paper and pencil tests.

The theoretical value of this research will enable educators to have additional reviews of ongoing research in the field of computer-aided testing and aid in their decisions on the acceptance and use of this test formats. The paper will also be of value to educators interested in crafting computer-aided tests or developers interested in converting paper and pencil tests to computer-aided format.

Initially this research was conducted to provide answers on the effect converting specific psychology tests from paper and pencil to computer-adapted versions. Traditionally psychological testing has been in personal interview or paper formatting. Before converting the test which was a simple operation there was a need for information and data to support the theory that computer-adapted test format would still be as reliable and valid for testing a undergraduate student population

subject pool as a paper and pencil test would be. It was imperative that the test be valid and also reliable over testing the student subject pool. In particular there were also issues of the computer interface and the effect the computer itself would have on test subjects if any. Current research by investigators of test anxiety have found computer anxiety to be at a minimal with undergraduate student pools as most are familiar with the use of a mouse or keyboard. Research has also found there is very little difference in the anxiety levels of students for more information on test anxiety refer to (Stedman, 1998) being tested using paper formats as oppose to that of computer-aided formats. The research overview conducted substantiates the conversion of paper and pencil to computer-aided testing format.

The references in the paper are intended to be a guide for obtaining more research on the use of computerized testing. Due to the constraint of the paper not all of the issues and risks of using computerized formats for testing have been covered. It is beneficial that the reader investigate additional sources of information to on the effect computerized tests can have on students as well as any biases that may effect minority populations and groups of students being tested without prior knowledge or skills using the computer interface (keyboard and mouse). It should be noted ETS states in its literature "You can take the test even if you have little or no computer skills" (ETS, 1999). Other potential problems of computerized test development may be due to cultural bias (Boodoo, 1998) or the cultural context of the test which is beyond the scope of this paper. Also relevant will be the need to take into consideration students with special needs (Braille, visually enhanced monitors or adjusted extended testing time), although these subjects are not within the scope of this paper they are a consideration for testing a select student populations, and are in some cases may be important enhancements allowing students the ability to interface computerized test formats.

The goal of the paper is to provide the reader current data that reflects the advantages of computer-aided tests over paper and pencil tests and the benefits to be gained by educators in using the computerized formats when administering exams. Educators are presented with examples and trends of computer-aided tests and are made aware of how selecting computer-aided test can be a good choice when testing student populations. In addition this paper provides information on the current trend of large testing organizations like Educational Testing Services, Sylvan Learning Centers as well as smaller companies that test for licensing purposes to move away from traditional paper and pencil testing and adopt computerized tests. The trend of using computer-aided testing over pencil and paper is shown through research to have benefits for educators in testing student knowledge.

### 3. CONCLUSION

Electronic Versus Paper-Based Testing? While the majority of test takers still have the opportunity to select from computerized or pencil-and-paper exams, that decision may be short lived as testing services move, albeit slowly, toward computer adaptive tests. The GRE is slated for complete computerization next year. The DAT and TOEFL are automated. Professional college tests, such as the LSAT and MCAT, are more cautiously moving in that direction. The TOEFL issues described in the literature ([http://www.nafsa.org/publications/ie/fall98\\_winter99/ru bin.html](http://www.nafsa.org/publications/ie/fall98_winter99/ru bin.html)) suggest that these are not easy changes. In particular, the switch from paper and pencil to computer for the three decade-old TOEFL exam was no simple matter. True, even critics of computer-based testing, or CBT, agree the new TOEFL has something for everyone. For test takers in densely populated areas, the CBT will be offered year-round, a vast improvement over the monthly (or even less frequent) offerings of the paper and pencil version. The computer TOEFL, delivered at Sylvan test centers, will provide a more uniform and comfortable test environment. No more scrambling for front-row seats to get a better shot at hearing the tape-recorded questions: students will have their own, volume-adjustable headsets. No more listening to English-language dialogue out of a black box: test takers will respond to questions while viewing a context-enhancing computer screen showing American college students speaking everyday English. No more questions that are too hard or too easy: through a computer-adaptive technique, the computer will adjust questions to the performance level of the test taker. For institutions that rely on TOEFL in making admission decisions, the CBT will likely better reflect candidates' ability to use English on a college campus, and scores will be received faster than in the past (2 weeks versus 6 weeks).

Other questions remain. Users of TOEFL scores and teachers of English are asking how to interpret the CBT TOEFL scores—which use a different scale than the old test—and how to prepare students for the new demands of a computer-based English test. Foreign student advisers from less developed areas of the world where computer literacy can't be taken for granted wonder whether students there will be disadvantaged relative to their computer-savvy counterparts. For the American institutions that rely on TOEFL scores to help determine a foreign student's enrollment eligibility, the new computer exam creates a few process issues with which they are only now starting to grapple. For one, the computer-based TOEFL is scored on a different scale than the paper and pencil version—on a range of 0-300 rather than 310-677. (The paper-based TOEFL originally was scored from 200 to 677; to avoid overlap with the CBT TOEFL; that range was truncated to 310-677 as of July 1998.) The new scoring involves a more dramatic shift than the "recentering" that William Paver,

assistant dean of graduate studies at the University of Texas, Austin, says the College Board did when it changed the total SAT score from 900 to 1000 a few years ago.

CATs are the wave of the future since they challenge students on an individual level. Questions are fielded one at a time, beginning with one of medium difficulty. Depending upon the response, follow-up questions become easier, harder or stay the same. Even if you guess, subsequent answers lead back to the appropriate level. The computer scores each question before selecting the next. Because fewer test questions provide more precise information about a student's abilities, computerization may create customized exams in the future. The major drawback for those experienced with standardized tests is the one-question-at-a-time format, which prevents skipping around. You must answer the question on the screen to move ahead. One of the biggest advantages, however, is immediate scoring. Official scores are mailed much more timely--trimming the wait from six weeks to two. The GRE CAT and GMAT allow cancellation prior to viewing scores. There are, however, restrictions on how soon you may try again.

#### 4. REFERENCES

- Bagin, C. B. & Rudner, L. M. (1993) "What Should Parents Know About Standardized Testing in Schools ERIC Clearinghouse on Assessment and Evaluation: U.S. Department of Education. Contract No. RR92024001.
- Boodoo, G. (1998). Addressing cultural context in the development of performance-based assessments and computer-adaptive testing: Preliminary validity considerations. Journal of Negro Education. Vol 67(3), 211-219.
- Brin, D. W. (1999, Spring). Sylvan Delves into Higher Education with Purchase of University in Spain. Wall Street Journal. Section B, 3.
- Fenske, R (Ph D.) Arizona State University. <http://tikun.ed.asu.edu/coe/coe/faculty/FENSE/papers/gradjs.htm> (1996).
- GRE Graduate Record Examinations. (1999) 1998-1999 Information and Registration Bulletin. ETS Educational Testing Service [Brochure]. ETS:
- Ji, C (1998). Predictive validity of the Graduate Record Examination in education. Psychological Reports. Vol 82 (3, Pt 1).
- Luect, R., & Nungester, R. (1998). Some practical examples of computer-adaptive sequential testing. Journal of Educational measurement. Vol. 35 (3), 229-249.
- McBride, J. R. (1998). Innovations in computer-based ability testing: Promise, problems, and perils. Hakel, M. D. In Beyond multiple choice: Evaluating alternatives to traditional testing for selection. (pp. 23-39). Mahwah, NJ: Erlbaum.
- Meijer, R. R. & Nering, M. L. Palmer (1999). Computerizing adaptive testing: Overview and introduction. Applied Psychological Measurement. Vol. 23(3), 187-194.
- Microsoft Inc., "Adaptive Testing Certification and Skills Assessment Group" [http://www.overlink.ru/mcse/FAQ/faq2\\_dop.htm](http://www.overlink.ru/mcse/FAQ/faq2_dop.htm) (1999).
- Mills, C. N. (1999). Development and introduction of a computer adaptive Graduate Record Examination General Test. In Drasgow, F (Ed) & Olson-Buchanan, J. B. (Ed); et al. Innovations in computerized assessment (pp. 117-135). Mahwah, NJ: Erlbaum.
- Russell, M. and Haney, W. (1997) "Testing Writing on Computers: An Experiment Comparing Student Performance on Tests Conducted via Computer and via Paper-and-Pencil." Education Policy Analysis Archives, (5:3).
- Saggio, J. J. "The Validity of Standardized Testing" American College Student (HED 591)
- Steadman, M. F. (1998). The effects of the ability to change answers on computer-based test performance of associate degree nursing students. Dissertation Abstracts International, Vol. 58(10-A) (Georgia State University, USA) No. AAM9812149.
- Vispoel, W. P. & Rocklin, T. & Wang, T. (1994). Individual differences and test administration procedures: a comparison of fixed-item, computerized-adaptive, self-adapted testing. Applied Measurement in education. Vol. 7(1), 53-79.
- Weisband, S. and Keisler, S. (1996) "Self Disclosure on Computer Forms: Meta-Analysis and Implications."
- Wenzel, R. A. (1999) the effects of ability to change answers on computer-based test performance of associate degree nursing students. Dissertation Abstracts International, Vol. 59 (10-A). (University of Michigan, USA)
- Wise, S. L. (1999) Understanding self-adapted testing: The perceived control hypothesis. Applied Measurement in Education. Vol. 7 (1), 15-24.