



# INFORMATION CAPSULE

## Research Services

Vol. 0918  
June 2010

Christie Blazer, Supervisor

### Computer-Based Assessments

#### *At a Glance*

*This Information Capsule reviews research conducted on computer-based assessments. Advantages and disadvantages associated with computer-based testing programs are summarized and research on the comparability of computer-based and paper-and-pencil assessments is reviewed. Overall, studies suggest that for most students, there are few if any performance differences when multiple-choice tests are taken on computer versus on paper. Research indicates, however, that the mode in which students take a test may have an impact on their performance when demographic characteristics, computer skills, computer type, test characteristics, item type, and content area tested are considered. In particular, some studies suggest that students with more computer skills receive higher scores on computer-based tests. Research has also found that open-ended items may lead to more differences in performance between computer-based and paper-and-pencil tests and that students' performance may decline when they are required to scroll through information on the computer screen in order to respond to questions. Researchers therefore suggest that the transition from paper-and-pencil to computerized tests be made very cautiously. This report also includes a brief summary of the statewide computer-based testing programs being implemented in Miami-Dade County Public Schools.*

Although paper-and-pencil tests remain the standard form of test-taking in U.S. schools, many states are exploring ways to convert their statewide assessment systems to computers. The popularity of computer-based testing has increased as schools' technology capabilities continue to grow and students become more comfortable using computers for various educational tasks (Education Commission of the States, 2010; Bridgeman, 2009; Texas Education Agency, 2008; Rabinowitz & Brandt, 2001). Tucker (2009) reported that approximately half of U.S. states use computers to deliver at least a portion of their annual state assessments, but noted that "even the most technologically advanced states have done little except replace the conventional paper-based, multiple-choice, fill-in-the-bubble tests with computerized versions of the same." Computer-based testing is already used successfully for military training exams, job application exams in the private sector, state drivers' license exams, entrance exams into postsecondary educational institutions, and certification exams by professional groups (Bridgeman, 2009; Clariana & Wallace, 2002).

#### **Advantages of Computer-Based Assessments**

Advocates of computer-based assessments maintain that they offer significant advantages over paper-and-pencil tests, including:

- Computer-based tests are capable of including more interactive and engaging question types, such as simulations, on-line experiments, and graphing, allowing for the measurement of skills not easily assessed by traditional paper-and-pencil tests. In addition, proponents of computerized tests argue that they are a better match with the way students are accustomed to learning (Csapó et al., 2010; Bridgeman, 2009; Buško, 2009; Kikis-Papadakis & Kollias, 2009; Kozma, 2009; Kyllonen, 2009; Lee, 2009; Martin, 2009; Scheuermann & Björnsson, 2009; Thompson & Weiss, 2009; Tucker, 2009).
- Computer-based tests can be adapted to individual students' ability levels. Computer-adaptive tests adjust item difficulty based on students' responses to previous items. Incorrect responses evoke less difficult items, while correct responses evoke increasingly difficult items. This results in a more refined profile of skill levels for each student (Education Commission of the States, 2010; Kozma, 2009; Moe, 2009; Scheuermann & Björnsson, 2009; van Lent, 2009; Gamire & Pearson, 2006).
- Computer-based assessments allow educators to collect data on students' testing strategies, intermediate progress, amount of time spent on each question, and thought processes, in addition to their final answers. This information is based on analyses of times and sequences in data records that track students' path through each task, their choices of which materials to access, and decisions about when to begin responding to items (Csapó et al., 2010; Bridgeman, 2009; Buško, 2009; Kozma, 2009; Martin, 2009; Thompson & Weiss, 2009; Tucker, 2009).
- Computer-based assessments can be more easily designed to meet the needs of special populations, including students with disabilities and those from diverse linguistic backgrounds (Gamire & Pearson, 2006).
- Quicker scoring of tests provides timely feedback to inform future instruction (Education Commission of the States, 2010; Kikis-Papadakis & Kollias, 2009; Kyllonen, 2009; van Lent, 2009; Puhan et al., 2007; Gamire & Pearson, 2006; Paek, 2005; Bennett, 2003).
- Computerized administrations result in greater standardization of test administrations. For example, computers manage test timing very accurately (Bridgeman, 2009).
- Additional educational tools can be made available on an item-specific basis. For example, dictionaries can be made available for certain questions and turned off for others; one part of a test might require a full scientific calculator while another part might require only a simple four-function calculator (Bridgeman, 2009).
- Computer-based tests provide several security advantages. Instead of storing testing materials at school sites for days before a test administration, tests can be sent over the Internet at the last minute, reducing the possibility of questions being exposed prior to the test. In addition, item sequences can be randomly scrambled for each student. When adaptive tests are used, students respond to different subsets of items so there is not one specific set of test questions that can be copied and distributed (Bridgeman, 2009; Buško, 2009; Moe, 2009; Thompson & Weiss, 2009).
- Electronic delivery is less expensive than printing and mailing large quantities of testing materials. In addition, errors found in test booklets or answer sheets can be quickly and easily corrected, instead of reprinting and reshipping testing materials at considerable expense (Bridgeman, 2009; van Lent, 2009; Bennett, 2003; Choi & Tinkler, 2002).
- Upon completion of the test, answer sheets and test booklets do not have to be mailed back to a central location for scoring, eliminating the chance that materials will be lost or damaged (Bridgeman, 2009; Rabinowitz & Brandt, 2001).

- Computer-based assessments reduce the costs associated with entering, collecting, aggregating, verifying, and analyzing data (Buško, 2009; Kozma, 2009).
- Computerized tests reduce teachers' assessment demands in the classroom. Staff time is reduced because there is no longer the need to process vast amounts of paper (Johnson & Green, 2004; Rabinowitz & Brandt, 2001).
- Computer-based tests significantly reduce the consumption of paper (Kikis-Papadakis & Kollias, 2009; Puhan et al., 2007; Paek, 2005).
- Most studies have reported that students prefer computer-based assessments over paper-and-pencil tests. However, it should be noted that correlations between enjoyment of computer-based tests and achievement have been found to be weak. In other words, students' preference for taking tests on computers doesn't necessarily translate into higher test scores (Education Commission of the States, 2010; Buško, 2009; Lee, 2009; Martin, 2009; Wang & Shin, 2009; Florida Department of Education, 2006; Higgins et al., 2005; Paek, 2005).

### **Disadvantages Associated with Computer-Based Assessments**

Efforts to computerize assessments have been hindered by a number of methodological and technological challenges. Disadvantages associated with computer-based assessments include:

- Computer crashes are more difficult to resolve than broken pencils. There is the potential that an entire testing session, along with all students' responses, could be lost. Back-up procedures are essential, both in terms of storing student responses and having alternative means to administer the test (Education Commission of the States, 2010; Bridgeman, 2009; Rabinowitz & Brandt, 2001). Kyllonen (2009) stated: "Computers add an extra layer of complication, require extra reviews, advanced set-ups, and tryouts."
- There are significant start-up costs for assessment systems that have previously been implemented only in paper and pencil format, including hardware, software, and network purchases, connectivity, item banking, staff training, and technical support (Education Commission of the States, 2010; Kikis-Papadakis & Kollias, 2009; Kozma, 2009; Kyllonen, 2009; Lee, 2009; Gamire & Pearson, 2006; Bennett, 2003).
- Computer-based assessments can lead to equity issues if some students have more access to computers and greater computer literacy skills than others. Research suggests that students with more computer skills perform at higher levels on computer-based tests than students with lower levels of computer skills (Csapó et al., 2010; Education Commission of the States, 2010; Thompson & Weiss, 2009; Gamire & Pearson, 2006; Paek, 2005; Poggio et al., 2005).
- Security concerns associated with computer-based tests center around staggered administrations of the same assessment (Kozma, 2009; van Lent, 2009; Bennett, 2003). In addition, Rabinowitz and Brandt (2001) noted that "a simple push of a button could send 'secure' test forms literally around the world." They concluded that states need to create multiple forms of each test, which will require the development of much larger item banks than most states currently have available.
- School computing facilities vary considerably and it is often difficult to ensure that students are provided with uniform testing environments. Equipment often varies from one school to the next and sometimes from one machine to the next within the same school. Variability in testing conditions and procedures, such as Internet connection speeds and hardware and software specifications, must be addressed (Csapó et al., 2010; Kikis-Papadakis & Kollias, 2009). Bennett (2003) suggested that equipment variations be controlled by establishing hardware and software standards, directly

manipulating resolution and font characteristics through the test delivery software, and designing items so that they display adequately at the lowest likely resolution.

- When large numbers of students take an assessment simultaneously, issues of scale must be addressed, such as network and server congestion, fluctuations in speed, and possible disruptions in service (Kozma, 2009; Kyllonen, 2009; Thompson & Weiss, 2009).
- Many schools lack the technical support needed to keep computerized systems functioning properly and equipment running smoothly (Education Commission of the States, 2010; Buško, 2009; Bennett, 2003).
- Most schools don't have the capacity to test all students on computers in one session. Therefore, administration of computer-based assessments usually involves significant changes to existing teaching schedules, as well as room, student, and personnel assignments. States, districts, and schools must decide how many testing sessions are needed, how many and which students will test during each session, and the specific dates and times of the testing window (Buško, 2009; Kozma, 2009; van Lent, 2009; Rabinowitz & Brandt, 2001).
- Considerable numbers of staff need to be trained in the administration of computerized tests. Test administrators need knowledge related to loading and/or accessing files, ensuring uniform assessment conditions, disabling software features (such as grammar checker for a writing test), and storing and transmitting files (Buško, 2009; Kikis-Papadakis & Kollias, 2009; Lee, 2009). Rabinowitz & Brandt (2001) noted that states must not underestimate the amount of staff training that is required in the early years of new programs.
- Scoring interactive design problems with open-ended responses is much more difficult than developing an answer key for multiple-choice questions (Bridgeman, 2009).

### **The Cost of Introducing Computer-Based Assessments**

Policymakers continue to debate whether computer-based assessments will increase or reduce test administration costs and no definitive conclusion has been reached. In some studies, the higher cost of computer-based tests, compared to paper-and-pencil tests, has been reported as an obstacle to the development of computer-based assessments. Increased costs include hardware, software, and network purchases, connectivity, item banking, staff training, and technical support (Education Commission of the States, 2010; Kikis-Papadakis & Kollias, 2009; Kozma, 2009; Kyllonen, 2009; Lee, 2009; Gamire & Pearson, 2006; Bennett, 2003). Other studies, however, suggest that computer-based assessment reduces the costs associated with the delivery and return of testing materials and the entering, collecting, aggregating, verifying, and analyzing of data (Bridgeman, 2009; Buško, 2009; Kozma, 2009; van Lent, 2009; Bennett, 2003; Choi & Tinkler, 2002; Rabinowitz & Brandt, 2001).

Farcot and Latour (2009) concluded: "The diversity of computer-based assessment is so large that searching for a unique, general, and transposable answer concerning the cost efficiency of CBA [computer based assessment] as opposed to P&P [pencil and paper assessment] is misleading. On the contrary, deciding between different CBA scenarios and P&P scenarios should be scrutinized on a case-by-case basis."

### **Comparability of Computer-Based and Paper-and-Pencil Assessments**

Educators need to ensure that a test presented on the computer measures the same knowledge and skills as its paper-and-pencil counterpart and that scores from computer-based test administrations have the same meaning as scores from paper-and-pencil administrations. Test scores should be

dependent on students' ability, not on the test administration mode. Furthermore, no student should be disadvantaged because of a change in test administration medium (Kikis-Papadakis & Kollias, 2009; Kozma, 2009; Martin & Binkley, 2009; Schroeders, 2009; van Lent, 2009; Paek, 2005; Poggio et al., 2005; Rabinowitz & Brandt, 2001)

The Florida Department of Education (2006) stated: "Choosing between computer-administered and paper-administered tests would be easier if there were clear, incontrovertible evidence that for all students there is no difference in results whether a test is taken on computer or by printed test materials."

Some studies suggest that students do not obtain the same results when they take an identical test on both computer and on paper. This finding is referred to as a "test mode effect." The test mode effect is the observation that performance tests measuring similar knowledge and skills yield different results when they are administered on computers versus with paper and pencil. For state and national assessments, comparability across delivery modes is important because assessments are usually offered on both computer and paper, since most schools don't have the infrastructure and equipment to test all of their students by computer. In these cases, scores from the two modes should be interchangeable. Comparability is also important when there is a transition from paper-and-pencil to computer-based delivery and educators want to compare students' performance across time (Csapó et al., 2010; Texas Education Agency, 2008; Crusoe, 2005; Bennett, 2003; Clariana & Wallace, 2002).

### **Research Comparing Performance on Computer and Paper Tests**

Overall, research on the comparability of computerized and paper-and-pencil assessments suggests that mode of administration has very little effect on students' performance (Moe, 2009; Schroeders, 2009; Sórenson & Andersen, 2009; Bennett et al., 2008; Wang et al., 2007; Horkay et al., 2005; Poggio et al., 2005). Paek's (2005) summary of comparability studies found that out of 97 cases, the results of computer-based and paper-and-pencil tests were comparable in 74 cases; in eight cases, the computer-based test appeared to be more difficult; and in 15 cases, the paper-and-pencil test appeared to be more difficult. The Texas Education Agency (2008) noted, however, that even a small effect can have significant consequences. For example, the Agency pointed out that a mode difference of even one point on a test can result in a substantial number of students not passing because they took the test in a different mode. Several studies have also found that even when overall test score differences between the two modes of administration are not significant, certain items may be more affected by mode of administration than others (Kim & Huynh, 2007; Higgins et al., 2005; Johnson & Green, 2004; Choi & Tinkler, 2002).

The following sections of this report review research conducted to determine if the mode in which students take a test has an impact on their performance, based on their demographic characteristics and computer skills; type of computer used to take the test; test characteristics; item type; and content area tested. The reader should note that several of the reviewed studies did not randomly assign students to testing conditions or control for their prior levels of academic achievement. In these cases, it cannot be stated with confidence that performance differences were due to testing mode. Higher scores may, in fact, have been caused by factors such as preexisting differences in students' levels of knowledge or computer literacy.

### **Research Comparing Performance on Computer and Paper Tests, Based on Students' Demographic Characteristics**

Some studies have found that, regardless of gender, students perform at similar levels when they take tests on computers versus on paper (Florida Department of Education, 2006; Paek, 2005; Poggio et al., 2005; Sim & Horton, 2005). On the other hand, a number of studies have found that boys outperform girls when tested on the computer, while girls perform significantly better on paper-and-pencil tests

(Csapó et al., 2009; Halldórsson et al., 2009; Lee, 2009; Martin & Binkley, 2009; Sórenson & Andersen, 2009; Higgins et al., 2005). Researchers have hypothesized several reasons for this finding. Some suggest that although gender gaps in volume of computer usage have closed rapidly over the last few years, boys are much more likely to play online games and use game-type software that are similar to the flash animations and video footage used with many computer-based test items. These activities expose boys more frequently to the content that appears in computerized tests. Others theorize that boys' higher performance on computerized tests may partially be explained by computer-based tests' lower reading load or a bias toward boys in the content of items included on computerized tests (Halldórsson et al., 2009; Martin & Binkley, 2009; Sórenson & Andersen, 2009; Crusoe, 2005).

Horkay and colleagues (2005) used the National Assessment of Educational Progress' (NAEP) Writing Online (WOL) study to examine differences in students' performance on computer-based and paper-and-pencil tests, based on their gender, ethnicity, parents' education level, income level (based on eligibility for free or reduced price lunch), and school location. WOL groups were composed of nationally representative groups of eighth grade students drawn from the main NAEP assessments. The researchers found no significant differences in either boys' or girls' performance on computer-based versus paper-and-pencil tests. Similarly, no significant differences were found between the scores of students who completed their tests on paper versus computers based on ethnicity (Asian/Pacific Islander, Black, Hispanic, White, or "Other"), parents' education level, and income level. The only time the researchers found significant differences in scores, based on test administration mode, was when performance was analyzed by school location. Students from urban fringe/large town locations were found to perform significantly higher on paper-and-pencil tests compared to computerized tests. No significant administration mode differences were found, however, for students from central city or rural/small town locations.

Poggio and colleagues (2005) compared students taking computer-based and paper-and-pencil equated forms of Kansas' state math assessment. Schools were given the opportunity to voluntarily administer the state's grade 7 math assessment on computers. The researchers found no differences in students' performance in the two administration modes, based on their income level (using eligibility for free and reduced price lunch as a measure of low-income status).

### **Research Comparing Performance on Computer and Paper Tests, Based on Students' Computer Skills**

Several researchers have noted that the replacement of paper-and-pencil assessments with computer-based assessments introduces equity issues into the testing situation. In the U.S., for example, surveys conducted for Pew Research Center's Internet & American Life Project in 2009 found that only 35 percent of low-income Americans (household income reported at \$20,000 or less) stated that they had broadband connections, while 85 percent of upper-income Americans (household incomes reported at over \$75,000) stated that they had access to these services (Horrigan, 2009). It is therefore possible that higher-income students have more familiarity and experience with computers.

Studies indicate that students with more computer skills receive higher scores on computer-based tests than students with fewer computer skills. Conversely, students with fewer computer skills and those who don't use computers on a regular basis have been found to perform better on paper-and-pencil tests (Csapó et al., 2010; Education Commission of the States, 2010; Bridgeman, 2009; Kyllonen, 2009; Gamire & Pearson, 2006; Paek, 2005; Poggio et al., 2005). Researchers have concluded, therefore, that computer-based assessments may place an unfair disadvantage on certain subgroups of learners who don't have as much opportunity to practice on the computer and become familiar with testing conditions (Kikis-Papadakis & Kollias, 2009; Rabinowitz & Brandt, 2001).

Russell, Goldberg, and O'Connor (2003) postulated that tests administered via a single medium, either computer or paper, underestimate the performance of students accustomed to working in the alternate medium. The researchers concluded that computer-based tests underestimate the performance of students who are not accustomed to working on computers because their inability to use the keyboard proficiently interferes with their ability to communicate in writing. On the other hand, paper-and-pencil tests underestimate the performance of students who are accustomed to working on computers because writing their answers on paper interferes with their ability to record and edit their ideas.

Bennett and colleagues (2008) analyzed results from the National Assessment of Educational Progress' 2001 Math Online (MOL) study. A nationally representative sample of eighth grade students was administered a computer-based math test and a test of computer facility (measuring computer experience, input accuracy, and input speed). In addition, a comparison group of students was administered a paper-and-pencil test containing the same math items. Students were randomly assigned to testing conditions. Results showed that students with greater input speed and accuracy received higher MOL scores. The researchers concluded that some students may have received higher scores than their equally mathematically proficient peers simply because of their more advanced computer skills.

Horkay and colleagues' (2005) nationally representative sample of eighth grade students participating in the National Assessment of Educational Progress' Writing Online (WOL) study found that students reporting more computer familiarity scored higher on the computer-based test than those reporting less computer familiarity. Computer familiarity added about 11 percentage points over paper-and-pencil writing scores to the prediction of performance.

A few studies have found no evidence that students with less computer experience score lower on computer-based assessments (Wang & Shin, 2009; Florida Department of Education, 2006; Paek, 2005). Higgins and colleagues' (2005) comparison of Vermont students randomly assigned to complete a reading comprehension test on computer or paper-and-pencil found no significant differences in test scores based on students' self-reported levels of computer fluidity (ability to use the mouse and keyboard) or computer literacy (familiarity with computing terms and functionality). However, they found that students with lower computer fluidity and/or literacy tended to receive the lowest scores.

### **Research Comparing Performance on Computer and Paper Tests, Based on Computer Characteristics**

Some researchers have concluded that computer hardware characteristics, such as monitor resolution, screen size, and responsiveness, affect student test scores (Csapó et al., 2010; Buško, 2009; Kyllonen, 2009; Wang & Shin, 2009; Florida Department of Education, 2006; Bennett, 2003). Monitor size can affect student performance because smaller monitors display the same amount of information as larger monitors, but the information looks smaller, making it harder to read the same question. Similarly, screen resolution affects the size of the text and how much information is shown (Csapó et al., 2010; Bennett, 2002). In addition, slower computer response times during online testing can have a negative impact on student performance (Florida Department of Education, 2006; Bennett, 2003).

Bridgeman, Lennon, and Jackenthal (2001) examined the effect of variations in screen size, resolution, and item presentation latency (the amount of time between answering one question and the presentation of the next question) on the performance of eleventh grade students completing SAT I verbal and math items on computers. Students were randomly assigned to computers with smaller or larger monitor sizes, lower or higher screen resolution, and no latency versus a five-second time delay. In math, the researchers found that monitor size and resolution had no impact on students' scores. Surprisingly, students with delayed item presentation had higher mean scores than students completing test items

without the delay. On verbal items, screen resolution, but not screen size, had a significant impact on test scores. Verbal scores were higher for students using high resolution displays.

As part of Horkay and colleagues' (2005) National Assessment of Educational Progress (NAEP) Writing Online (WOL) study, the researchers examined the effect of computer type on writing performance. Results from their analyses were inconsistent. In the first of two analyses, 88 students were randomly assigned to complete the test on either a desktop computer or a laptop computer. Those who took the test on laptops scored significantly lower than students taking the test on desktop computers, but only on one of two essays. In the second analysis, conducted on the larger WOL main NAEP sample (1,313 students), no significant score differences were found between students taking the WOL on laptops versus desktop computers. It should be noted that results within the larger sample differed by gender. While there was no difference between the scores for male students on laptops versus desktop computers, female students scored significantly higher on desktop computers than on laptops.

### **Research Comparing Performance on Computer and Paper Tests, Based on Test Characteristics**

In general, most experts suggest that the more complicated it is to present or take a test on a computer, the greater the possibility of mode effects. Scores obtained from computer-based and paper-and-pencil tests have been found to be equivalent when the computer-based test is constructed to look similar to the paper-and-pencil version of the test (Poggio et al., 2005; Pommerich, 2004; Russell et al., 2003).

Several studies have suggested that mode effects are insignificant when all of the information for an item is presented entirely on the computer screen, but that students' performance declines when they are required to scroll through information on the computer screen in order to answer questions (Texas Education Agency, 2008; Paek, 2005; Pommerich, 2004). Choi and Tinkler (2002) converted paper-and-pencil reading and math subtests from Oregon's statewide achievement test to computerized versions of the tests. Classrooms within schools were randomly assigned to computer-based or paper-and-pencil testing conditions. The researchers found that scrolling reading passages on computer screens interfered with students' test-taking behavior, especially for younger (grade 3) students. The researchers recommended that students be provided with page-up and page-down buttons in place of a vertical scroll and be allowed to use an electronic marker to highlight passages in order to reduce the negative impact of scrolling on test-taking.

Bridgeman, Lennon, and Jackenthal (2001) asked grade 11 students to evaluate the extent to which the following five features interfered with taking a computer-based test: font size, use of a mouse, screen clarity, screen size, and scrolling. The only feature rated as interfering for the majority of students was scrolling, with 66 percent reporting that it interfered with test-taking to some degree. In contrast, less than 25 percent of respondents indicated that font size, screen clarity or screen size interfered with test-taking (39 percent reported that use of a mouse interfered with test-taking). Further analysis of responses found that students who had been randomly assigned to take the test on smaller monitors were more likely to report that scrolling interfered "a great deal" with their test-taking. In actuality, the amount of scrolling necessary was identical for both monitor sizes (15 inch and 17 inch), but was perceived as a greater hindrance when students took the test using a smaller monitor.

Higgins, Russell, and Hoffmann (2005) examined the test scores of fourth grade students who were randomly assigned to complete the same computer-based reading comprehension test in one of three modes: on paper; on computer with scrolling reading passages; or on computer with passages divided into sections that were presented as whole pages of text. They found that students completing the test on paper received the highest mean score, followed by the whole page group, and then by the scrolling group, although there were no significant differences among the scores of the three groups.



The researchers concluded: “Overall, students were neither advantaged or disadvantaged by the mode of test delivery.”

Halldórsson and colleagues (2009) analyzed students’ scores from the 2006 Programme for International Student Assessment (PISA) Computer-Based Assessment (CBA) of science. CBA items were categorized in terms of their degree of interactivity (for example, whether the item required specific skills such as dragging and dropping or whether it was a relatively simple item that involved watching a video and clicking in a response box). In the three countries that participated in the CBA (Denmark, Iceland, and Korea), high interactivity items were found to be more difficult for students than low interactivity items.

Researchers have also concluded that students’ inability to review and/or skip individual test items has a negative effect on computer-based test performance. Basic features available to students during paper-and-pencil tests that are not always available to students taking tests on computers include the ability to skip items and answer them later in the test; the ability to review items already answered; and the ability to change answers to items (Johnson & Green, 2006; Russell et al., 2003). Johnson and Green (2006) concluded that students taking the test in paper-and-pencil format “possessed a degree of independence and control on paper that allowed them access to strategies that could facilitate their performance.”

Pommerich (2004) conducted two comparability studies on a sample of eleventh and twelfth grade students in 40 schools. She concluded that students were sensitive to how test items were presented. For example, students taking computer-based tests were better able to focus on some items because relevant sections of those items were centered in item windows and students were not distracted by extraneous information. However, Pommerich hypothesized that students taking paper-and-pencil tests were more likely to experience “positional memory,” whereby they remembered the location of information given in the passage, based on its spatial location both on the page and within the document. Conversely, some students taking computer-based tests had difficulty locating information in passages because scrolling allowed only relative spatial orientation. Some students testing on computers also had difficulty comparing information across tables or figures that did not appear on screen simultaneously.

Two studies found that more transcription errors were made when students attempted to transfer problems from the computer screen to scratch paper space, and then back to the computer screen (Johnson & Green, 2006; Russell et al., 2003). The researchers suggested that tests incorporate methods that allow students to make notes on screen in order to minimize transfer errors.

For timed tests, computer-based tests may give students an advantage over their peers taking tests in the traditional paper-and-pencil format (Wang & Shin, 2009; Paek, 2005; Poggio et al., 2005). Pomplun, Frey, and Becker (2002) reported that students taking a speeded test scored about seven points higher when tests were administered on computers. Pommerich (2004) found that the difference in performance across modes tended to be greater for items at the end of the test (when students were running short on time) than at the beginning of the test. Researchers have attributed this test mode effect to the speed of using a mouse rather than a separate answer sheet to record responses (Pommerich, 2004; Pomplun et al., 2002).

### **Research Comparing Performance on Computer and Paper Tests, Based on Item Type**

Most research suggests that student performance on multiple-choice tests does not differ significantly based on mode of administration. However, comparability studies conducted on open-ended test questions have produced mixed results (Education Commission of the States, 2010; Wang & Shin,

2009; Wang et al., 2007; Florida Department of Education, 2006; Paek, 2005). Bennett (2003) theorized that the threat to comparability is greater with open-ended items because they make heavier demands on students' computer skills than multiple-choice questions and risk measuring computer proficiency rather than subject area knowledge.

Bennett and colleagues (2008) analyzed results from the National Assessment of Educational Progress' 2001 Math Online (MOL) study. A nationally representative sample of eighth grade students from 110 U.S. schools was administered a computer-based math test or a paper-and-pencil test containing the same math items. Students were randomly assigned to computer-based or paper-and-pencil testing conditions. The researchers found that open-ended items required more adaptation than multiple-choice items in order to be rendered on screen and that open-ended items became more difficult when they were presented on computers as compared to paper. The researchers concluded that when moving paper test items to the computer, it may be harder to hold difficulty constant for open-ended items than for multiple-choice items.

Keng, McClarty, and Davis (2006) reported that the test mode effect was significantly larger for Texas Assessment of Knowledge and Skills English/language arts items with long passages and for math items involving graphing and geometric manipulations that required more scrolling through the screen. Csapó and colleagues (2009) found the largest differences in computer-based and paper-and-pencil inductive reasoning test scores when students completed open-ended items requiring mathematical calculations.

The Texas Education Agency (2008) conducted a literature review of studies that analyzed the comparability of computer-based and paper-and-pencil open-ended items, including short written text, numeric responses, and geometric drawings. The literature review concluded that two factors led to differential performance on open-ended items. One was students' degree of familiarity with typing text responses on the computer. All studies found that students with computer experience, whether through instruction or tests previously taken, tended to perform as well, if not better on computer-based versions of the open-ended items. The second factor was the feasibility or amount of work required to convert the paper-and-pencil version of an open-ended item to the computer version. Larger performance differences were found when tests required more adaptation in order to be presented on screen. The studies cited by the Texas Education Agency also found that computer-based essay performance was higher than paper essay performance when computerized tests included standard word processing supports (such as cut, copy, paste, and undo) and a way for students to monitor how much they had written compared to the amount of space they had been provided.

Several studies found that students who took essay tests on computers wrote longer essays and received higher scores than randomly assigned groups of students taking the same tests on paper (MacCann et al., 2002; Russell & Plati, 2000; Russell & Haney, 1997; Wolfe et al., 1996). In contrast, Horkay and colleagues' (2005) nationwide sample of eighth grade students participating in the National Assessment of Educational Progress' Writing Online (WOL) study found no significant difference between the number of words students wrote on computerized tests, compared to the number of words they wrote on paper-and-pencil tests.

On a related note, another type of comparability concerns scoring differences when open-ended responses are submitted to raters in handwritten versus typewritten form. In this case, the question is not how delivery mode affects students' performance, but how the format of the response affects the scorer's judgment of that performance (van Lent, 2009; Bennett, 2003).

There is no consensus on whether typed and handwritten essays are judged differently. Some studies have found no significant differences between the ratings that identical typed and handwritten essays receive. Other studies have shown that typed essays are more likely to be judged by expert raters as being

proficient. Still other studies have reported that typed essays are scored more reliably because any bias that exists due to handwriting quality is eliminated (Vantage Learning, 2006). Bennett (2003) recommended that raters be trained to avoid scoring biases that may be associated with response format.

Russell and Tao (2004) studied how response format affected raters' judgments. They analyzed fourth, eighth, and tenth grade students' essays from the Massachusetts Comprehensive Assessment System Language Arts Test. All essays were handwritten by students and subsequently typed on computer by the researchers. The essays were presented to raters in one of three formats: handwritten; printed in single-spaced 12-point text; and printed in double-spaced 14-point text. The handwritten versions were found to receive significantly higher scores than the typed versions, but no differences were found between the two forms of typed essays. As part of their study, Russell and Tao also asked raters to identify errors in 80 eighth grade essays. Half of the essays appeared in handwritten form and half appeared in print. Error categories included spelling, punctuation, capitalization, awkward transitions, and confusing phrases or sentences. Results indicated that the raters detected significantly more spelling errors and more confusing phrases or sentences when essays were presented in print. Russell and Tao suggested that errors may stand out more when responses are typed because they are easier to read and because raters' expectations for typed responses may be higher.

### **Research Comparing Performance on Computer and Paper Tests, Based on Content Area Tested**

Content area can have an impact on the comparability of alternative test versions, although the few studies conducted on the effects of test administration mode across various academic subject areas have produced inconclusive findings (Wang & Shin, 2009; Kim & Huynh, 2007).

The Texas Education Agency (2008) reviewed studies conducted on the comparability of computer-based and paper-and-pencil tests in state testing programs across the country. In the subject areas of math, language arts, science, and social studies, most studies indicated that tests were comparable in overall difficulty across the two administration modes. Comparability studies conducted in Texas for the Texas Assessment of Knowledge and Skills (TAKS) program yielded mixed results, but tended to suggest mode effects in math and reading/English language arts.

Kingston's (2009) meta-analysis of K-12 multiple-choice tests found that computer administration provided a small advantage for language arts and social studies tests, while math tests favored paper-and-pencil tests. Similarly, the Education Commission of the States (2010) reported that computerized math exams seemed to pose the most trouble for students.

Pommerich's (2004) comparability studies randomly assigned students to computer-based and paper-and-pencil administrations of English language, reading, and science reasoning tests. She found that average student performance was higher on computer for English language tests and higher on paper for reading tests. Results from administrations of the science reasoning tests were inconclusive.

Russell (1999) analyzed the performance of students randomly assigned to take tests on computers or in the traditional paper-and-pencil format, using open-ended items from the National Assessment of Educational Progress and the Massachusetts Comprehensive Assessment System. No significant differences were found between the language arts and math scores of students who took the test on computer and those who took the test on paper. In science, however, students taking the test on computer significantly outperformed students taking the test on paper.

The Texas Education Agency (2008) also examined whether there were differences in student performance, based on mode of administration, at the individual item level. Several studies identified characteristics of items that may have led to differences in student performance. In math, these were hypothesized to be items that:

- were multi-part or required a lot of scrolling to see the entire item;
- were intended to determine how effectively students can manipulate a physical tool, such as a ruler or protractor;
- required students to create drawings, enter a lengthy amount of text, or produce mathematical formulas;
- required graphing or geometric manipulations;
- required extended tutorials or lengthy item-specific directions for responding;
- required paper stimulus materials; and
- required significant changes to the formatting in order to be presented on computers.

In reading, items that required on-screen scrolling or paging because of lengthy passages or item content appeared to lead to differences in student performance depending on mode of test administration. Features such as the layout of passages, location of line breaks, ease of navigation through items, alignment of items with reading passages, and capacity to highlight relevant text all contributed to item-level performance differences. Providing students with computer-based tools, such as highlighters and review markers, tended to help the performance of students testing on computers.

In science, allowing students to view only graphics relevant to a specific item on screen, without the distraction of extraneous information from other items, helped the performance of students taking tests on computers. Science items requiring scrolling on screen favored students taking paper-and-pencil tests. In social studies, the research reviewed by the Texas Education Agency (2008) did not identify any observable item characteristics that may have led to performance differences.

### **On a Local Note**

The Florida Department of Education (FLDOE) is moving toward computer-based testing in the state's public schools. In March 2010, the FLDOE gave all high schools, alternative education centers, and adult education schools the option of administering the FCAT Retake on computers instead of paper. In Miami-Dade County Public Schools (M-DCPS), five schools elected to administer the Spring 2010 FCAT Retake on computers. A total of 144 students participated in the computerized FCAT Reading Retake and 59 students participated in the FCAT Mathematics Retake. Shortly after the administration of the computer-based FCAT Retake, Florida's Chancellor of Public Schools notified districts that schools that had participated in the computer-based administration of the FCAT Retake would be given the option of taking the test again on paper. The Chancellor stated: "While many schools and students have reported positive testing experiences, we realize that others encountered technical issues during testing due to glitches in the test platform provided by our contractor that may have been distracting." Problems encountered during the computerized administration of the FCAT Retake included lag times of as much as 40 minutes to sign on to the practice test site; students being unable to log on to the test site in their first attempt and test chairpersons having to manually resume testing; and non-functioning calculator applications. Four of the five M-DCPS schools that had administered the computerized FCAT Retake re-administered the test on paper. The FLDOE will credit students with the higher of the two scores.

Administration of the Florida Assessments for Instruction and Reading (FAIR) is overseen by staff from the Division of Language Arts/Reading in all of the District's schools. The assessment is Internet-based. In kindergarten and grades 1 and 2, the test is administered to students individually by teachers and teachers enter data into computers. In grades 3-12, students enter their own responses, and the assessment can be administered individually or in a computer lab setting. The FLDOE offers FAIR free of charge to provide teachers with screening, progress monitoring, and diagnostic information on each student and identify those most likely to be on or above grade level in reading by the end of the school year. M-DCPS tests all students in kindergarten and grades 1 and 2 using the FAIR and all low-performing (FCAT Achievement Levels 1 and 2) students in grades 3-12. FAIR is administered three times per year (fall, winter, and spring).

In May 2010, The Florida Department of Education field tested the Algebra I End-of-Course test in selected schools throughout the state. The FLDOE selected 12 M-DCPS senior high schools, representing 3,181 students, to participate in the field test. Students completed the test on computers and the District was given a five-day window in which to complete the testing. Students were administered a subset of items, not the full End-of-Course test, and no scores will be released for students or schools. Difficulties encountered during the Algebra End-of-Course field test included problems logging on to the TestNav testing platform; students kicked out of the system before logging on, or being unable to log on to the test site in their first attempt, with test chairpersons having to manually resume testing; and non-functioning applications (including calculators and highlighting capabilities). In at least three cases, students took the entire test but the computer recorded no responses for them.

During the 2010-11 school year, all of the District's middle and senior high schools will be impacted by computer-based testing. The FCAT Mathematics Retake will be administered on computers at all high schools, alternative education centers, and adult education schools in Fall 2010 and Spring 2011, with the option of administering the Reading Retake on computers as well. The FCAT Mathematics High School Graduation Test (Grade 10) will be administered on computers in Spring 2011. In May 2011, the FLDOE will conduct a full administration of the baseline Algebra I End-of-Course test. All senior high schools, alternative education centers, middle schools, and K-8 centers that offer Algebra I will be required to administer the End-of-Course test. Finally, the FLDOE will field test End-of-Course tests in Geometry and Biology in May 2011.

### **Summary**

The popularity of computer-based testing has increased in schools across the country, with approximately half of U.S. states using computers to deliver at least a portion of their annual state assessments. Advantages of computer-based tests include their use of more interactive and engaging questions; the faster turnaround time between test-taking and receipt of scores to guide instruction; and lower costs associated with the delivery and return of materials and analysis of data. Disadvantages associated with computer-based assessments include complications resulting from computer crashes; minimizing variations in equipment between and within schools to standardize students' testing experience; and providing schools with the required technical support and staff training to keep the testing process running smoothly.

Studies suggest that for most students, there is very little difference in test scores when multiple-choice tests are administered on computers versus with paper-and-pencil. However, research indicates that students' demographic characteristics and computer skills, as well as computer and test characteristics, item type, and content area tested may all play a role in the comparability of computer-based and paper-and-pencil tests. Although results from the studies reviewed in this report were mixed, some general patterns emerged. The studies suggest that students with more computer skills score higher on computer-based tests than students with fewer skills, indicating that some computer-based tests are measuring students' computer proficiency rather than, or in addition to, their content area skills. In addition, research indicates that open-ended items lead to more performance differentials between computer-based and paper-and-pencil tests, compared to tests comprised of multiple-choice items. Some studies suggest that computer characteristics (such as screen size and monitor resolution) may affect students' performance on computer-based tests. Research has found that scrolling in particular may lead to declines in performance when students are required to scroll through information on the computer screen in order to respond to items. Finally, some studies have found that boys tend to outperform girls on computer-based tests, while girls tend to outperform boys on paper-and-pencil tests.

This report also summarized statewide computer-based testing programs being implemented in M-DCPS. During the 2009-10 school year, the Florida Department of Education gave all high schools, alternative education centers, and adult education schools the option of administering the FCAT Retake on computers instead of paper. In addition, the Florida Assessments for Instruction and Reading were administered on computers at all M-DCPS schools. In May 2010, the state's Algebra I End-of-Course test will be field tested on computers in 12 M-DCPS schools. During the 2010-11 school year, all M-DCPS middle and senior high schools will be impacted by computer-based testing. The FCAT Mathematics Retake, FCAT Mathematics High School Graduation Test, baseline administration of the Algebra I End-of-Course test, and field tests of the Geometry and Biology End-of-Course tests will all be administered on computers.

Based on the inconsistency of research results and the many still unanswered questions surrounding the comparability of computer-based and paper-and-pencil tests, Lee (2009) suggested that the transition from paper-and-pencil to computer-based tests be made cautiously. Csapó and associates (2009) recommended that computer-based assessments be piloted in a few schools to ensure that they produce equivalent results before they are administered on a statewide or districtwide basis. Bennett (2002) suggested piloting multiple-choice tests in a small number of subjects in a few grade levels, gradually scaling up to more schools, more subjects, more grades, as well as to tests that include open-ended items. Researchers also urge that separate comparability studies be conducted for each new assessment program, based on its own unique tests and technology. Test scores from computer-based and paper-and-pencil test administrations should be equated or, in the most extreme cases, separate scales may need to be created for computer versus paper versions of tests (Texas Education Agency, 2008; Pommerich, 2004; Bennett, 2003).

All reports distributed by Research Services can be accessed at <http://drs.dadeschools.net>.

## References

- Bennett, R.E. (2002). Using Electronic Assessment to Measure Student Performance. *National Governors Association Center for Best Practices Issue Brief*. Retrieved from <http://www.nga.org/cda/files/ELECTRONICASSESSMENT.pdf>.
- Bennett, R.E. (2003). *Online Assessment and the Comparability of Score Meaning*. Educational Testing Service, Princeton, NJ. Retrieved from <http://www.ets.org/Media/Research/pdf/RM-03-05-Bennett.pdf>.
- Bennett, R.E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., & Yan, F. (2008). Does It Matter if I Take My Mathematics Test on Computer? A Second Empirical Study of Mode Effects in NAEP. *The Journal of Technology, Learning, and Assessment*, 6(9). Retrieved from <http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1140&context=jtla>.
- Bridgeman, B. (2009). Experiences from Large-Scale Computer-Based Testing in the USA. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.
- Bridgeman, B., Lennon, M.L., & Jackenthal, A. (2001). Effects of Screen Size, Screen Resolution, and Display Rate on Computer-Based Testing Performance. *Educational Testing Service Research Report RR-01-23*. Retrieved from <http://www.ets.org/Media/Research/pdf/RR-01-23-Bridgeman.pdf>.

- Buško, V. (2009). Shifting from Paper-and-Pencil to Computer-Based Testing: Requisites, Challenges and Consequences for Testing Outcomes. A Croatian Perspective. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.
- Choi, S.W., & Tinkler, T. (2002). *Evaluating Comparability of Paper-and-Pencil and Computer-Based Assessment in a K-12 Setting*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA, April 2002. Retrieved from <http://www.ncme.org/repository/incoming/100.pdf>.
- Clariana, R.B., & Wallace, P.E. (2002). Paper-Based Versus Computer-Based Assessment: Key Factors Associated with the Test Mode Effect. *British Journal of Educational Technology*, 33(5), 593-602.
- Crusoe, D. (2005). *A Discussion of Gender Diversity in Computer-Based Assessment*. Retrieved from [http://www.bitculture.org/storage/DHC\\_Gender\\_Div\\_EdDRvw0705.pdf](http://www.bitculture.org/storage/DHC_Gender_Div_EdDRvw0705.pdf).
- Csapó, B., Molnár, G., & Tóth, K.R. (2009). Comparing Paper-and-Pencil and Online Assessment of Reasoning Skills: A Pilot Study for Introducing Electronic Testing in Large-Scale Assessment in Hungary. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.
- Csapó, B., Ainley, J., Bennett, R., Latour, T., & Law, N. (2010). *Draft White Paper 3: Technological Issues for Computer-Based Assessment*. Assessment and Teaching of 21<sup>st</sup> Century Skills, The University of Melbourne, Australia. Retrieved from <http://www.atc21s.org/GetAssets.axd?FilePath=/Assets/Files/dc7c5be7-0b3a-4b7d-8408-cc610800cc76.pdf>.
- Education Commission of the States. (2010). *Assessment: Computer-Based*. Retrieved from <http://www.ecs.org/html/issue.asp?issueID=12&subIssueID=76>.
- Farcot, M., & Latour, T. (2009). Transitioning to Computer-Based Assessments: A Question of Costs. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.
- Florida Department of Education. (2006). *What Do We Know About Choosing To Take a High-Stakes Test on a Computer?* Retrieved from <http://www.fldoe.org/asp/k12memo/pdf/WhatDoWeKnowAboutChoosingToTakeAHighStakesTestOnAComputer.pdf>.
- Gamire, E., & Pearson, G. (Eds.). (2006). *Tech Tally: Approaches to Assessing Technological Literacy*. Washington, DC: The National Academies Press.
- Halldórsson, A.M., McKelvie, P., & Björnsson, J.K. (2009). Are Icelandic Boys Really Better on Computerized Tests Than Conventional Ones? Interaction Between Gender, Test Modality and Test Performance. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.
- Higgins, J., Russell, M., & Hoffmann, T. (2005). Examining the Effect of Computer-Based Passage Presentation on Reading Test Performance. *The Journal of Technology, Learning, and Assessment*, 3(4). Retrieved from <http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1054&context=jtla>.
- Horkay, N., Bennett, R.E., Allen, N., & Kaplan, B. (2005). Part II: Online Assessment in Writing. In B. Sandene, N. Horkay, R.E. Bennett, N. Allen, J. Braswell, B. Kaplan, & A. Oranje (Eds.). *Online Assessment in*

*Mathematics and Writing: Reports from the NAEP Technology-Based Assessment Project* (NCES 2005-457). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Horrigan, J.B. (2009). *Home Broadband Adoption 2009*. Pew Internet and American Life Project. Retrieved from <http://www.pewinternet.org/~media/Files/Reports/2009/Home-Broadband-Adoption-2009.pdf>.

Johnson, M., & Green, S. (2004). *On-Line Assessment: The Impact of Mode on Student Performance*. Paper presented at the British Educational Research Association Annual Conference, Manchester, England, September 2004. Retrieved from [http://www.cambridgeassessment.org.uk/ca/digitalAssets/113933 On-line Assessment. The Impact of Mode on Student Performanc.pdf](http://www.cambridgeassessment.org.uk/ca/digitalAssets/113933_On-line_Assessment_The_Impact_of_Mode_on_Student_Performanc.pdf).

Johnson, M., & Green, S. (2006). On-Line Mathematics Assessment: The Impact of Mode of Performance and Question Answering Strategies. *Journal of Technology, Learning, and Assessment*, 4(5). ERIC Document Reproduction Service No. EJ843854.

Keng, L., McClarty, K.L., & Davis, L.L. (2006). Item-Level Comparative Analysis of Online and Paper Administrations of the Texas Assessment of Knowledge and Skills. *Applied Measurement in Education*, 21(3), 207-226.

Kikis-Papadakis, K., & Kollias, A. (2009). Reflections on Paper-and-Pencil Tests to eAssessments: Narrow and Broadband Paths to 21<sup>st</sup> Century Challenges. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.

Kim, D., & Huynh, H. (2007). Comparability of Computer and Paper-and-Pencil Versions of Algebra and Biology Assessments. *The Journal of Technology, Learning, and Assessment*, 6(4). Retrieved from <http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1111&context=jtla>.

Kingston, N.M. (2009). Comparability of Computer- and Paper-Administered Multiple-Choice Tests for K-12 Populations: A Synthesis. *Applied Measurement in Education*, 22(1), 22-37.

Kozma, R. (2009). Transforming Education: Assessing and Teaching 21<sup>st</sup> Century Skills. Assessment Call to Action. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.

Kyllonen, P.C. (2009). New Constructs, Methods, & Directions for Computer-Based Assessment. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.

Lee, M. (2009). CBAS in Korea: Experiences, Results and Challenges. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.

MacCann, R., Eastment, B., & Pickering, S. (2002). Responding to Free Response Examination Questions: Computer Versus Pen and Paper. *British Journal of Educational Technology*, 33(2), 173-188.

Martin, R. (2009). Utilising the Potential of Computer-Delivered Surveys in Assessing Scientific Literacy. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.



- Martin, R., & Binkley, M. (2009). Gender Differences in Cognitive Tests: A Consequence of Gender-Dependent Preferences for Specific Information Presentation Formats? In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.
- Moe, E. (2009). Introducing Large-Scale Computerised Assessment: Lessons Learned and Future Challenges. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.
- Paek, P. (2005). Recent Trends in Comparability Studies. *Pearson Educational Measurement Research Report 05-05*. Retrieved from [http://www.pemsolutions.com/downloads/research/TrendsCompStudies\\_rr0505.pdf](http://www.pemsolutions.com/downloads/research/TrendsCompStudies_rr0505.pdf).
- Poggio, J., Glasnapp, D.R., Yang, X., & Poggio, A.J. (2005). A Comparative Evaluation of Score Results from Computerized and Paper & Pencil Mathematics Testing in a Large Scale State Assessment Program. *The Journal of Technology, Learning, and Assessment*, 3(6). Retrieved from <http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1057&context=jtla>.
- Pommerich, M. (2004). Developing Computerized Versions of Paper-and-Pencil Tests: Mode Effects for Passage-Based Tests. *The Journal of Technology, Learning, and Assessment*, 2(6). Retrieved from <http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1002&context=jtla>.
- Pomplun, M., Frey, S., & Becker, D.F. (2002). The Score Equivalence of Paper and Computerized Versions of a Speeded Test of Reading Comprehension. *Educational and Psychological Measurement*, 62(2), 337-354.
- Puhan, G., Boughton, K., & Kim, S. (2007). Examining Differences in Examinee Performance in Paper and Pencil and Computerized Testing. *The Journal of Technology, Learning, and Assessment*, 6(3). Retrieved from <http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1115&context=jtla>.
- Rabinowitz, S., & Brandt, T. (2001). Computer-Based Assessment: Can It Deliver On Its Promise? *WestEd Knowledge Brief*. ERIC Document Reproduction Service No. ED462447.
- Russell, M. (1999). Testing on Computers: A Follow-Up Study Comparing Performance On Computer and On Paper. *Education Policy Analysis Archives*, 7(20). Retrieved from <http://epaa.asu.edu/ojs/article/view/555/678>.
- Russell, M., & 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). Retrieved from <http://epaa.asu.edu/ojs/article/view/604/726>.
- Russell, M., & Plati, T. (2000). *Effects of Computer Versus Paper Administration of a State-Mandated Writing Assessment*. Technology and Assessment Study Collaborative, The Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, Chestnut Hill, MA. Retrieved from <http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1012&context=intasc>.
- Russell, M., & Tao, W. (2004). Effects of Handwriting and Computer-Print on Composition Scores: A Follow-Up to Powers, Fowles, Farnum, & Ramsey. *Practical Assessment, Research & Evaluation*, 9(1). Retrieved from <http://pareonline.net/getvn.asp?v=9&n=1>.
- Russell, M., Goldberg, A., & O'Connor, K. (2003). Computer-Based Testing and Validity: A Look Back Into the Future. *Assessment in Education*, 10(3), 279-294.

- Scheuermann, F., & Björnsson, J. (Eds.). (2009). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.
- Schroeders, U. (2009). Testing for Equivalence of Test Data Across Media. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.
- Sim, G., & Horton, M. (2005). *Performance and Attitude of Children in Computer Based Versus Paper Based Testing*. Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2005. Chesapeake, VA: AACE.
- Sørensen, H., & Andersen, A.M. (2009). How Did Danish Students Solve the PISA CBAS Items? Right and Wrong Answers From a Gender Perspective. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.
- Texas Education Agency. (2008). *A Review of Literature on the Comparability of Scores Obtained from Examinees on Computer-Based and Paper-Based Tests*. Retrieved from [http://ritter.tea.state.tx.us/student.assessment/resources/techdigest/Technical Reports/2008\\_literature\\_review\\_of\\_comparability\\_report.pdf](http://ritter.tea.state.tx.us/student.assessment/resources/techdigest/Technical Reports/2008_literature_review_of_comparability_report.pdf).
- Thompson, N.A., & Weiss, D.J. (2009). Computerized and Adaptive Testing in Educational Assessment. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.
- Tucker, B. (2009). The Next Generation of Testing. *Educational Leadership*, November 2009, 48-53.
- van Lent, G. (2009). Risks and Benefits of CBT versus PBT in High-Stakes Testing: Introducing Key Concerns and Decision Making Aspects for Educational Authorities. In F. Scheuermann & J. Björnsson (Eds.). *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-Scale Testing*. Luxembourg: Office for Official Publications of the European Communities.
- Vantage Learning. (2006). *Research Summary: Effects of Mode of Delivery on Constructed Response and Selected Response Assessments*. Retrieved from [http://www.vantagelearning.com/docs/intellimetric/IM\\_ResearchSummary\\_Effects\\_of\\_Mode\\_of\\_Assessment\\_Delivery.pdf](http://www.vantagelearning.com/docs/intellimetric/IM_ResearchSummary_Effects_of_Mode_of_Assessment_Delivery.pdf).
- Wang, H., & Shin, C.D. (2009). Computer-Based & Paper-Pencil Test Comparability Studies. *Pearson Education Test, Measurement & Research Services Bulletin*, Issue 9. Retrieved from [http://www.pearsonassessments.com/NR/rdonlyres/93727FC9-96D3-4EA5-B807-5153EF17C431/0/Bulletin\\_9.pdf](http://www.pearsonassessments.com/NR/rdonlyres/93727FC9-96D3-4EA5-B807-5153EF17C431/0/Bulletin_9.pdf).
- Wang, S., Jiao, H., Young, M.J., Brooks, T., & Olson, J. (2007). A Meta-Analysis of Testing Mode Effects in Grades K-12 Mathematics Tests. *Educational and Psychological Measurement*, 67(2), 219-238.
- Wolfe, E.W., Bolton, S., Feltovich, B., & Niday, D.M. (1996). The Influence of Student Experience with Word Processors on the Quality of Essays Written for a Direct Writing Assessment. *Assessing Writing*, 3(2), 123-147.