Essays in Education

Volume 1 Article 3

Spring 2002

It's Time to Upgrade: Tests and Administration Procedures for the New Millennium

Michael Russell

National Board on Educational Testing and Public Policy Center for the Study of Testing, Evaluation and Educational Policy Boston College, MA

It's Time to Upgrade: Tests and Administration Procedures for the New Millennium

Michael Russell
National Board on Educational Testing and Public Policy
Center for the Study of Testing, Evaluation and Educational Policy
Boston College, MA

Abstract:

Increasing use of computers in schools has led to a mis-alignment between the way some students develop skill and knowledge and how they are tested. This paper reviews past research that demonstrates that paper-based tests that require students to produce written responses underestimate the achievement of students who are accustomed to writing on computer. The paper then explores how learning that occurs through other instructional uses of computers is not adequately captured by current testing practices. The paper argues that new approaches should be explored to better measure student learning.

A Tale of a Time Past

Imagine. Virtual reality transfers us back in time.[1] The eraser-capped pencil was invented just twenty years ago. Since then, it has matured from a novelty item flashed by the wealthy to a common household tool. Students across the country are replacing their quill and inkwells with pencils. When asked why, they exclaim, "It's so much easier to write with a pencil. I can let the ideas flow from my mind to paper without constantly dipping my quill in ink, blowing dry each row of ink before moving on to the next line, or worrying about how to correct mistakes." And in classrooms, teachers observe that students write more and better with pencils than they did with the old fashioned writing tool.

Yet, once a year, the state demands that students set aside their pencils and dig out their dusty inkwells and worn quills. For it is testing time and to insure standardization, all students must write with the same tool.

For students who have not yet transitioned to the pencil, test day is just like any other day, only the stakes are a little higher. For those students fortunate enough to attend schools that make pencils available to all students or who have parents that have invested in pencils for them, testing day is one filled with frustration. They have lost the art of quill dipping and have difficulty applying the proper amount of ink to their quill tips. Accustomed to recording ideas as they flow from their mind, they forget to blow dry each line of text. As the clock ticks, they find both their words on paper and thoughts

in mind growing more smudged. At the end of the day, the best writers still produce good work, but few pencil users are proud of their performances. And their teachers sense that their students' essays do not reflect their true achievements.

The Story of Today

Fast forward to the new millennium. Pencils are omnipresent, but are rapidly being replaced by computers. Increasing numbers of students across the country are gaining access to this new writing tool at school and at home. Beyond increasing the fluidity with which students record their thoughts, computers help students see errors in their writing, make more revisions to their writing, produce fewer spelling errors, and develop a better sense of audience (Dauite, 1984, 1985; Wresch, 1984; Elbow, 1981; Owston & Wideman, 1997). Students who use computers regularly see measurable improvements in the quality of their writing (Cochran-Smith, Paris & Khan, 1991, Owston, Murphy & Wideman, 1992; Owston & Wideman, 1997). Yet, on test day, computers are forbidden and the performance of students accustomed to writing with computers suffers. Teachers know their students' essays do not reflect their true achievements.

What Research Reveals

Several studies have shown that the mode of administration, that is paper versus computer, has little impact on students' performance on multiple-choice tests (Bunderson, Inouye & Olsen, 1989; Mead & Drasgow, 1993). More recent research, however, shows that young people who are accustomed to writing with computers perform significantly worse on open-ended (that is, not multiple choice) questions administered on paper as compared with the same questions administered via computer (Russell & Haney, 1997; Russell, 1999; Russell & Plati, 2000).

Research on this topic began with a puzzle. While evaluating the progress of student learning in the Accelerated Learning Laboratory (ALL), a high-tech school in Worcester, MA, teachers were surprised by the results from the second year of assessments. Since infusing the school with computers, the amount of writing students performed in school had increased sharply. Yet, student scores on writing tests declined significantly during the second year of the new program.

To help solve the puzzle, a randomized experiment was conducted, with one group of sixty-eight students taking math, science and language arts tests, including both multiple-choice and open-ended items, on paper, and another group of forty-six students taking the same tests on computer (but without access to word processing tools, such as spell-checking or grammar-checking). Before scoring, answers written by hand were transcribed to computer text so that raters could not distinguish them from those done on computer. There were two major findings. First, the multiple-choice test results did not differ much by mode of administration. Second, the results for the open-ended tests differed significantly by mode of administration. For the ALL School students who were accustomed to writing on the computer, responses written on computer were much better than those written by hand. This finding occurred across all three subjects tested and on both short answer and extended answer items. The effects were so large that when

students wrote on paper, only 30 percent performed at a "passing" level; when they wrote on computer, 67 percent "passed" (Russell & Haney, 1997).

Two years later, a more sophisticated study was conducted, this time using open-ended items from the new Massachusetts state test (the Massachusetts Comprehensive Assessment System or MCAS) and the National Assessment of Educational Progress (NAEP) in the areas of language arts, science and math. Again, eighth grade students from two middle schools in Worcester, MA were randomly assigned to groups. Within each subject area, each group was given the same test items, with one group answering on paper and the other on computer. In addition, data were collected on students' keyboarding speed and prior computer use. As in the first study, all answers written by hand were transcribed to computer text before scoring.

In the second study, which included about two hundred students, large differences between computer and paper-and-pencil administration were again evident on the language arts tests. For students who could keyboard moderately well (20 words per minute or more), performance on computer was much better than on paper. For these students, the difference between performance on computer and on paper was roughly a half standard deviation. According to test norms, this difference is larger than the amount students' scores typically change between grade 7 and grade 8 on standardized tests (Haney, Madaus, & Lyons, 1993). For the MCAS, this difference in performance could easily raise students' scores from the "failing" to the "passing" level (Russell, 1999).

In the second study, however, findings were not consistent across all levels of keyboarding proficiency. As keyboarding speed decreased, the benefit of computer administration became smaller. And at very low levels of keyboarding speed, taking the test on computer diminished students' performance. Similarly, taking the math test on computer had a negative effect on students' scores. This effect, however, became less pronounced as keyboarding speed increased.

A third study, conducted during the spring of 2000, found similar effects for students in grades four, eight and ten. In addition, this most recent study also found that students accustomed to writing with eMates (portable writing devices capable of displaying about twenty lines of text) also performed significantly worse when forced to perform a state writing test on paper. Furthermore, this study found that the mode of administration effect was about 1.5 times larger for eighth grade students with special education plans for language arts than for all other eighth grade students.

The effect was so large that eliminating the mode of administration effect for all five written items on the state language arts test would have a dramatic impact on district level results. As figure 1 indicates, based on 1999 MCAS results, 19 percent of the fourth graders classified as "Needs Improvement" would move up to the "Proficient" performance level. An additional 5 percent of students who were classified as "Proficient" would be deemed "Advanced." Similarly, figure 2 shows that in grade eight, four percent of students would move from the "Needs Improvement" category to the "Proficient" category and that 13 percent more students would be deemed "Advanced." And within one elementary school (Figure 3), the percentage of students performing at or above the "Proficient" level would nearly double from 39 percent to 67 percent.

Figure 1: Mode of Administration Effect on Grade 4 1999 MCAS Results

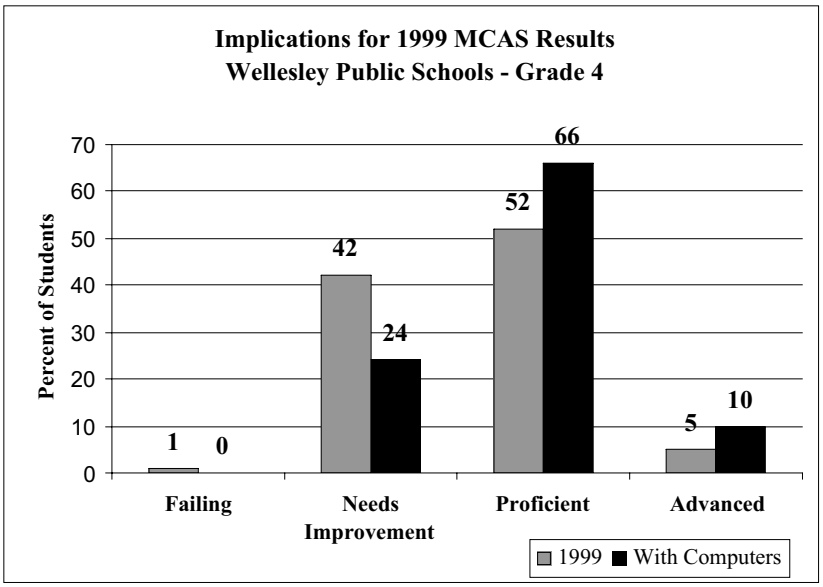


Figure 2: Mode of Administration Effect on Grade 8 1999 MCAS Results

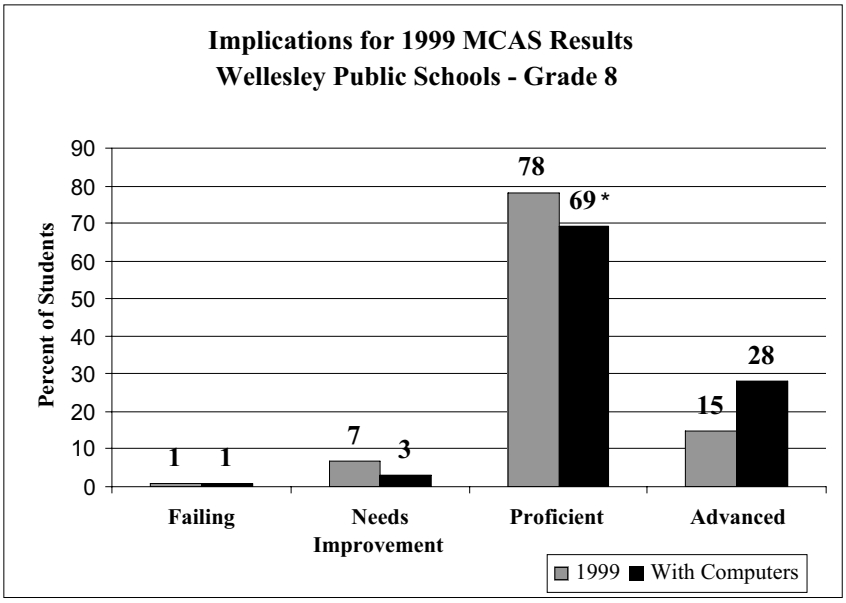
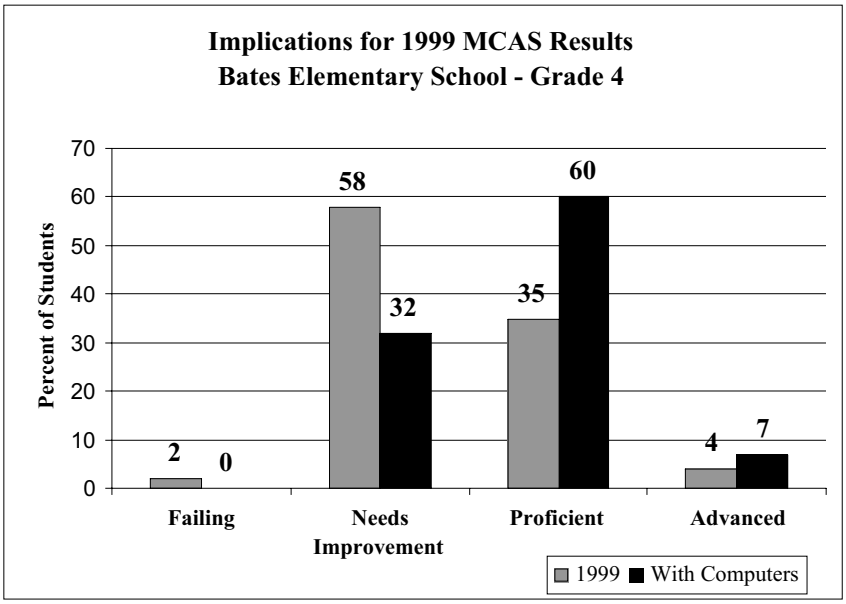


Figure 3: Mode of Administration Effect on Bates Elementary School 1999 MCAS Results



The Conflict

This mis-measurement of students accustomed to writing with computers is problematic for three reasons. First, several states make important high-stakes decisions about students and their schools based on these test scores. For students accustomed to writing with a computer, these decisions are based on tests that underestimate their achievement. Decisions based on these inaccurate test scores mischaracterize students and schools who have adopted computers as their tool for writing.

Second, the public and the press are increasingly pointing to stagnant test scores as an indication that investments in educational technologies are not impacting student learning. In the case of writing, the current paper-and-pencil tests underestimate the achievement of students who are using computers for writing. As a result, improvements may be masked by underestimated scores.

Third, as pressure to improve scores on state tests increases, some schools are beginning to limit computer use for writing so that students are not mis-measured by paper-and-pencil tests (see Russell, 1999). After reviewing the first study described above and following the introduction of the new paper-and-pencil MCAS test in Massachusetts, one school required students to write more on paper and less on computer (Russell, 1999). In another Massachusetts school system, the principal feared that students who wrote regularly on computer would lose penmanship skills, which might lead to lower scores on the new state test. This school increased penmanship instruction across all grades while also decreasing students' time on computers (Holmes, 1999). Such strategies, in effect reducing computer use in schools to better prepare students for low-tech tests, may be pragmatic given the high stakes attached to many state tests. But

they are also shortsighted in light of students' entry after graduation into an increasingly high-tech world and workplace.

A Short-Term Solution: Provide Options

One solution state testing programs might adopt to reduce the mode of administration effect is to allow students to select the mode in which open-ended responses are composed. For the past decade, the Province of Alberta has employed this strategy for its graduation-testing program. Over the past five years, the province has seen the percentage of students opting to perform the English, Social Studies, Biology, and French tests on computer increase from 6.7 percent in 1996 to 24.5 percent in 2000. Within high schools, the percentage of students opting to perform the test on a computer ranges from 0 to 80 percent (A. Sakyi, personal communication, April 26, 2000).

Although this approach adds to the complexity of test administration procedures (see Russell & Haney, 2000 for a fuller review of added complexities), providing students the option of working on either paper or computer would create writing conditions that more closely reflect normal practice. In turn, students would be better able to demonstrate their best possible work under the circumstances.

To date, however, state testing programs have expressed considerable resistance to providing students the option of writing on paper or on computer. Their reasons include:

1. Concern that penmanship skills might deteriorate if students did not have to write state tests by hand.
2. Concern that computers provide students with an unfair advantage because there is something about the computer rather than the student that improves the quality of student writing.
3. Concern that students in schools that do not have large numbers of computers would not be able to take advantage of a computer option.
4. Concern that a computer option might increase scores for students in wealthier districts that invest in computers and, in turn, increase differences in the scores between urban and suburban districts.

Although I do not have room here to respond more fully to these concerns, I believe they pale in importance given the high-stakes decisions that test scores are being used to make about students, schools, and the impact of technology on student learning. As Elana Scraba (personal communication, April 18, 2000), the Assistant Director of Alberta's Learning Assessment Branch, reasons, "We are interested in students being able to demonstrate their best possible work under the circumstances, and have always believed that their writing tools should be compatible with how they normally write." The less performance on tests reflects what students can actually do, the less we can say about student or school achievement and the role computers play in this achievement.

Writing: The Tip of the Iceberg

When used appropriately over an extended period of time, computers can have a positive impact on students' writing skills (Cochran-Smith, Paris & Khan, 1991; Owston, Murphy & Wideman, 1992; Owston & Wideman, 1997). But writing is just one of several types of learning that computers can help students develop. Some other areas of learning include: a) problem solving; b) research; c) non-linear thinking; d) developing a better understanding of concepts related to physics, earth science, biology, and chemistry; e) collaboration; f) spatial reasoning; g) statistics; h) media; I) music theory; and J) modeling and simulating complex mathematic, social, and scientific relationships. Among all of these areas, only one is extensively measured by current state-mandated testing programs, namely writing. Yet, as summarized above, there is mounting evidence that current testing methods seem to do a better job mis-measuring than measuring the impact of computers on writing.

While I sympathize with those students and educators who believe that there are already too many tests, I also believe that new instruments are needed to more accurately measure the skills and knowledge computers help students develop. To the public, tests are seductive because they seem to provide objective, scientific measures of what students know and can do. In reality, tests are not as objective and scientific as the public believes. However, as Hawkins (1996, p 49) wrote, tests and other methods of assessment do "provide us with the terms, images, and emotions of what it is important to know." Until tests that measure the types of learning enabled by computers are developed, it is likely that the public and policy makers will under-value the types of learning influenced by computers. In turn, the public and policy makers will continue to under-estimate the impact computers have on student learning.

Arguably, some current tests measure problem solving skills, scientific understanding, statistics and spatial relations. However, the number and types of items used to test students' achievement in these areas is insufficient for assessing the impact of computer use on these skills. As just one example, an evaluation of a Massachusetts school district recently conducted by Russell (2000) revealed that most third and fourth grade teachers in this district use computers as a part of their mathematics instruction to help students develop spatial reasoning skills. However, on the state's fourth grade test (MCAS), only two of the 39 released items relate to spatial reasoning. It would be tenuous, at best, to use changes in MCAS scores to examine the impact computer use has on students' math achievement.

Similarly, most mathematics tests include items that test students' mathematical problem solving skills. Typically, these items take the form of word problems for which students must define a function that represents the relationship described, plug in the appropriate numbers, and perform accurate computations. While it is important for students to develop these mathematical problem-solving skills, these skills are not what advocates of computer use envision when they discuss the potential impacts of computers on students' problem solving skills.

Problem solving with computers is more than just decoding text to define functions. As Dwyer (1996, p.18) describes, when developing problem-solving skills with

computers, "students are encouraged to critically assess data, to discover relationships and patterns, to compare and contrast, to transform information into something new." To help students assimilate, organize, and present their learning, some teachers have students use HyperCard and other multimedia tools.

After studying HyperCard use in a small set of ACOT classrooms, Tierney (1996, p.176) concluded: "Technology appears to have increased the likelihood of students' being able to pursue multiple lines of thought and entertain different perspectives. Ideas were no longer treated as unidimensional and sequential; the technology allowed students to embed ideas within other ideas, as well as pursue other forms of multilayering and interconnecting ideas. Students began spending a great deal of time considering layout, that is, how the issues that they were wrestling with might be explored across an array of still pictures, video segments, text segments, and sound clips."

These findings are echoed by teachers in other schools. After studying technology use across classrooms in one school district, Russell (2000, p.11) wrote: "In addition to exposing students to a larger body of information related to the topic of study, creating HyperStudio stacks also requires students to more carefully plan how they integrate and present this information. As one teacher explains, 'First they do the research and identify what it is they want to include in their stack. They then create a flow chart that depicts how the pieces fit together. They sketch their stack on paper and then begin putting it into the computer.' Through this process, students develop their planning skills and learn to anticipate how information will be received by their audience."

Despite the skill development enabled by HyperCard and other multimedia authoring tools, students who develop complex, high quality products using HyperCard do not necessarily perform well on current tests. While studying the impact of computers on student learning in the Apple Classrooms of Tomorrow project, Baker, Herman, and Gearhart (1996, p.198) found that "…a sizeable portion of students who used HyperCard well to express their understanding of principles, themes, facts, and relationships were so-so or worse performers judged by more traditional forms of testing." Over the past decade these and similar findings have led proponents of computer use in schools to conclude that technology enables students to develop new competencies, "some of which were not being captured by traditional assessment measures" (Fisher, Dwyer, & Yocam, 1996, p.5). While I support this conclusion, I also believe critics of computers in schools are beginning to see this argument as a well-worn cover for "lukewarm results" (Jane Healy as quoted by Westreich, 2000).

Upgrading Testing Methods

It is time that testing and accountability programs develop and apply instruments and testing procedures that capture the types of learning impacted by computer use. To make this happen, several steps are required.

First, educators and parents must demand that the way students are assessed matches the medium in which they typically work. Advocates for disabled students have long argued that state and local assessment programs should "allow students the same

assistance in the assessment process as they have in the learning process…" and reason that "it is only fair that the assessment of what they have learned should allow for them to demonstrate their knowledge and skills in the way most appropriate to them" (Hehir, 2000, p. 50). I believe that the same argument applies to all students. Students who are accustomed to writing with computers in school or at home should be allowed to write with computers while being tested. Similarly, as some testing programs have begun to allow, students who are accustomed to working with graphing or traditional calculators should be allowed to use these during tests (with the exception of tests that measure students' ability to perform calculations).

Second, educators and advocates of computer use in schools must insist that testing programs develop tests that measure students' technology skills. Despite the large investments schools have made in computer-related technologies, only two states collect information about students' technology skills. And, until recently, paper-based multiple-choice tests were employed in both states. Although teachers use computers to help students develop a wide variety of skills, a thorough examination of the impacts of computers on student learning must include measures of students' computer skills. As Westreich (1996, p.23) notes, what are often termed basic computer skills such as keyboarding are in fact "important skill[s] that one needs in order to take maximum advantage of technology." During the past decade, many observers have also touted these skills as essential for the workplace (see Smith, 1999 for a fuller review of this literature). For students who do not have access to computers at home, the development of these essential computer skills represents an important impact of computer use in schools.

Third, instruments that measure the "other types of learning" possible with computers must also be developed. But, before these instruments can be developed, educators and researchers must be clearer about what these new types of learning are. It is not enough to say that computers allow students to develop problem solving, simulation or modeling skills. Test development begins by defining the domain and constructs to be measured. Catch-phrases like problem solving, simulating, and modeling do not provide clear descriptions of a domain or construct. As an example, researchers at the Educational Testing Service are currently developing a computer-aided assessment task that will test students' simulation skills. During a review of a preliminary version of this task, questions were raised as to whether the task was intended to test a student's ability to develop a simulation of a scientific experiment or use a simulator to simulate a scientific experiment. Although computers are used in schools to develop both types of skills, there is a clear difference between the two types of skills. To assist test developers, descriptions of these "new types of learning" must become more precise.

Finally, we must anticipate that computer-related technology will continue to evolve much faster than the technology of testing. Although I cannot predict what tomorrow's computer-related technologies will look like or what types of learning they will enable, we must narrow the gap between emerging computer-related technologies and new testing methods. The problem is similar to the disjuncture between the research community's findings about technology and what teachers have learned from these studies. To remedy the gap between research and practice, Norris, Smolka and Soloway (1999) recommend that researchers collaborate with teachers, curriculum developers, psychologists, and other professionals who work with students to find out what information

is truly useful for educators. Similarly, I argue that researchers must work more closely with test developers, developers of new educational technologies, teachers, cognitive scientists, and students to predict how these new technologies might affect student learning and to develop instruments that measure these constructs before these new technologies have permeated large numbers of schools.

McNabb, Hawkes & Rouk (1999, p.4) are correct: "Standardized test scores have become the accepted measure with which policymakers and the public gauge the benefits of educational investments." Acknowledging this as reality, educators and researchers must be proactive in establishing administration procedures and instruments that provide more accurate measures of the types of learning educational technology is believed to impact. Until these instruments and procedures are penned, testing programs will forever be mis-measuring the impact of computers on student learning.

References

Baker, E., Herman, J., & Gearhart, M. (1996). Does technology work in school? Why evaluation cannot tell the full story. In C. Fisher, D. Dwyer, and K. Yocam (Eds.), Education and technology: Reflections on computing in classrooms. San Francisco: Apple Press.

Bunderson, C., Inouye, D., & Olsen, J. (1989). The four generations of computerized educational measurement. In Linn, R. L. (Ed.), Educational measurement (3rd ed.). Washington, D.C.: American Council on Education.

Cochran-Smith, M, Paris, C., & Kahn, J. (1991). Learning to write differently: Beginning writers and word processing. Norwood, NJ: Ablex.

Daiute, C. (1984). Can the computer stimulate writers' inner dialogues? In W. Wresch (Ed.), The computer in composition instruction: A writer's tool. Urbana, IL: National Council of Teachers of English.

Daiute, C. (1985). Writing and computers. Don Mills, Ontario: Addison-Wesley.

Dwyer, D. (1996). Learning in the age of technology. In C. Fisher, D. Dwyer, and K. Yocam (Eds.), Education and technology: Reflections on computing in classrooms. San Francisco: Apple Press.

Elbow, P. (1981). Writing with power: Techniques for mastering the writing process. New York: Oxford University Press.

Fisher, C., Dwyer, D. and Yocam, K. (1996). Education and technology: Reflections on computing in classrooms. San Francisco: Apple Press.

Haney, W., Madaus, G., & Lyons, R. (1993). The fractured marketplace for standardized testing. Boston, MA: Kluwer Academic Publishers.

Hawkins, J. (1996). Dilemmas. In C. Fisher, D. Dwyer, and K. Yocam (Eds.), Education and technology: Reflections on computing in classrooms. San Francisco: Apple Press.

Hehir, T. (2000). Some assessments treat learning-disabled students unfairly. In D. Gordon (Ed.), The digital classroom. Cambridge, MA: Harvard Education Letter.

Holmes, R. (1999). A gender bias in the MCAS? MetroWest Town Online. Retrieved August 26, 2000, from http://www.townonline.com/metrowest/archive/022499/.

McNabb, M., Hawkes, M. & Rouk, U. (1999). Critical issues in evaluating the effectiveness of technology. Report prepared for the Secretary's Conference on Educational Technology: Evaluating the Effectiveness of Technology, Washington, DC. Retrieved December 15, 2001, from http://www.ed.gov/Technology/TechCong/1999/confsum.html.

Mead, A. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. Psychological Bulletin, 114, 449-58.

Norris, C., Smolka, J., and Soloway, E. (1999). Convergent analysis: A method for extracting the value from research studies on technology in education. Paper prepared for the Secretary's Conference on Educational Technology, U.S. Department of Education, Washington, DC, July 1999. Retrieved December 15, 2001, from http://www.ed.gov/Technology/TechConf/1999/whitepapers/paper2.html.

Owston, R. & Wideman, H. (1997). Word processors and children's writing in a high-computer-access setting. Journal of Research on Computing in Education, 30, 202-220.

Owston, R., Murphy, S., & Wideman, H. (1992). The effects of word processing on students writing ability. Research in the Teaching of English, 26, 249-276.

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 December 15, 2001, from http://olam.ed.asu.edu/epaa/v5n3.html.

Russell, M. & Haney, W. (2000). Bridging the Gap Between Testing and Technology in Schools. Education Policy Analysis Archives, 8(19). Retrieved December 15, 2001, from http://epaa.asu.edu/epaa/v8n19.html.

Russell, M. & Plati, T. (2000). Mode of Administration Effects on MCAS Composition Performance for Grades Four, Eight and Ten. Report prepared for the Massachusetts Department of Education. Retrieved December 15, 2001, from http://www.nbetpp.bc.edu/statements/ws052200.pdf.

Russell, M. (1999). Testing writing on computers: A follow-up study comparing performance on computer and on paper. Educational Policy Analysis Archives, 7(20). Retrieved December 15, 2001, from http://epaa.asu.edu/epaa/v7n20/.

Russell, M. (2000). Wellesley Public Schools curriculum technology program evaluation report. Wellesley, MA: Wellesley Public Schools.

Smith. J. (1999). Tracking the mathematics of automobile production: Are schools failing to prepare students for work? American Educational Research Journal, 36(4), 835-878.

Tierney, R. (1996). Redefining computer appropriation: A five-year study of ACOT students. In C. Fisher, D. Dwyer, and K. Yocam (Eds.), Education and technology: Reflections on computing in Classrooms. San Francisco: Apple Press.

Westreich, J. (2000). High-Tech kids: Trailblazers or guinea pigs? In D. Gordon. Cambridge, MA: Harvard Education Letter.

Wresch, W. (1984). The computer in composition instruction: A writer's tool. Urbana, IL: National Council of Teachers of English.

[1] The Tale of a Time Past is not historically accurate and is presented as a fictional analogy for a present-day problem.