Research Summary

Effects of Mode of Delivery on
Constructed Response and Selected Response Assessments

Vantage Learning

110 Terry Drive Suite 100
Newtown, PA 18940
www.vantagelearning.com

Research Summary

Effects of Mode of Delivery for
Constructed Response and Selected Response Assessments

Computers are fast becoming more and more vital to both learning and assessment. Most students have had at least some experience using a computer to compose writing and learn basic computer skills. As technology advances, a few important questions arise: If students are assessed using a computer instead of traditional paper-and-pencil methods, will the results differ? When students write using a computer, is there a difference in the quality or quantity of the writing? Is something gained or lost when a student composes an essay using a computer versus handwriting it? Also, on a standardized multiple-choice assessment, do students perform at the same level whether the assessment is presented via computer or paper-and-pencil? The purpose of this document is to detail and summarize existing research on the possibility of mode differences in constructive and selective response assessments, in both individuals and group populations.

Constructed Response

At first glance, composing writing on a computer versus using the traditional paper-and-pencil method can appear to be quite different. Many people can type and handwrite at different speeds. Revision using a computer involves a few strokes of the keyboard, whereas revision using paper-and-pencil can require considerably more time. The critical issue is twofold: Do students write better in one mode compared to the other? Do raters judge writing differently based on the mode in which it was written?

There are an abundance of studies that compare the differences between typed and handwritten essays. However, the research is very mixed on whether any mode differences truly exist. Many studies indicate that essays composed on computers are of a higher quality than those that are handwritten. Typed essays tend to be longer, more formal, and contain fewer mechanical errors (Bangert-Drowns 1989; Nichols 1996; Russell & Haney 1996; Wolfe, Bolton, Feltovich, & Niday 1996). When typing, students are more likely to organize their thoughts into paragraphs (Russell & Haney 1996). They also revise their work more often, and fewer errors remain in their final draft (Grejda & Hannafin 1992). Other studies indicate that no such differences exist; typed and handwritten essays are essentially equivalent in length and form (Harrington, Shermis, & Rollins 2000; Puhan & Boughton 2004; Wolfe & Manalo 2004). Still other studies conclude that when students use the pencil-and-paper method to compose essays, they tend to be of higher quality. These essays tend to have a stronger overall voice than typed essays (Powers, Fowles, Farnum, & Ramsay 1994; Craig 2001). The research is certainly mixed with regards to the effect of mode of delivery for constructive response items.

There is also no consensus on whether typed or handwritten essays are judged differently. Some studies indicate that when students type their essays, the writers are more likely to be judged by expert raters as being proficient. Typing essays allows some students to better display their true writing ability (Russell & Haney 1996). There is also an indication that essays composed on computers are scored more reliably. Because the text of every essay appears identical, any bias that exists due to handwriting quality is eliminated (Wolfe & Manalo 2004; Bridgeman & Cooper 1998). Other studies show that no such differences exist, with typed and handwritten essays being essentially equivalent. When typed essays are transcribed, both forms tend to be judged equally (Harrington, Shermis, & Rollins 2000; Nichols 1996; Puhan & Boughton 2004; NAEP 2005). Finally, some studies suggest that raters tend to be more lenient on handwritten essays. In a comparison of typed and transcribed essays, the typed essays appear shorter, with errors that were much more visible (Wolfe et al. 1996). Raters tend to have higher expectations of typed text (Craig 2001; Wolfe & Manalo 2004). When raters have difficulty determining what is written, they often "give the benefit of the doubt" to handwritten essays (Puhan & Boughton 2004). Other raters have indicated that they sometimes mark down an essay with the handwriting is so poor due to an association that poor handwriting and poor quality of writing are interrelated. These different research findings regarding rater judgment in conjunction with mode of delivery suggest that there are many factors at play, and there may not be one clear answer.

Any mode differences that do exist may be better explained on a more individual, student-by-student basis. Research indicates that most students prefer to use a computer to compose writing rather than paper-and-pencil. These students find this writing process to be more enjoyable and sociable (Bangert-Drowns 1989). However, students' computer and typing skills vary widely. There is a general consensus that the level of comfort and amount of experience students have with using a computer has a significant effect on the quality of their writing. If students are accustomed to writing on a computer, they tend to produce higher-quality writing when they type their essays (Russell & Haney 1997). Even in studies that indicated that there were no mode differences, it was determined that in the case of specific individuals, more hands-on computer skill allowed for better performance (NAEP 2005). This finding could have serious implications. Since most statewide tests are still completed using paper-and-pencil, "the abilities of millions of students used to writing on a computer could be severely underestimated" (Russell & Haney 1997). Other studies indicate that while computer experience does not have an effect on the quality of writing, students that lack computer experience are at a disadvantage when typing their essays (Manalo & Wolfe 2000; Wolfe et al 1996; Puhan & Boughton 2004; Bridgeman & Cooper 1998). It appears that "the use of word-processors by examinees with weaker computer and keyboarding skills interferes with the production of writing, but no such interference is encountered by examinees with stronger computer skills because keyboarding has become an automated process for these examinees" (Wolfe & Manalo 2004). The research supports the notion that as long as a student has some computer skills, testing via computer allows them to effectively compose writing that is either equivalent or superior to writing that is composed using pencil-and-paper.

Selected Response

The research on effects of mode of delivery has certainly not been limited to constructed response items. There are also concerns regarding the effect of mode of delivery for selected response (multiple-choice) tests. If the same test is presented to students in two modes (pencil-and-paper and computer), will students perform better when the test is administered in one mode over the other?

Some studies indicate that students perform better when multiple-choice tests are administered using the computer. Students are often able to complete the assessment in less time (Russell & Haney 1996). When the two forms are identical, students receive higher scores when the test is presented on the computer (Parshall & Kromrey 1993). Other studies indicate that mode differences did not exist (Peek 2005), while others suggest that the same questions are perceived as more difficult when presented via computer (NAEP 2005; Choi & Tinkler 2002). However, much of the differences in findings can be related to the equivalence of the two modes. If computer-based tests closely mirror those presented in paper-and-pencil format, the results should be comparable. Giving students the capability to see the entire question at the same time, and review and revise answers as they would on a paper test, allows for similar test-taking strategies across modes. "Evidence has accumulated to the point where it appears that in many traditional multiple-choice test settings, the computer may be used as a medium to administer tests without any significant effect on student performance" (Peek 2005). When users are able to effectively interact with a computer, they are likely to achieve similar results on both computer and paper tests.

As with constructed response assessments, any mode differences on multiple choice assessments appear on an individual basis. Although most students prefer using the computer, the scores of students often vary depending on the mode of the assessment (Parshall & Kromrey 1993; Choi & Tinkler 2002). The Math Online Study, conducted by the NAEP, suggests that performance on computer-based test items depends on how familiar a student is with using a computer. Students who are more familiar with the computer, and have better typing skills, are more likely to perform better on the computer-based test. "This finding suggests that computer familiarity may distort the measurement of mathematics achievement when tests are administered online to students who lack basic technology skills" (NAEP 2005). These results are mirrored by several other studies (Parshall & Kromrey 1993; Choi & Tinkler 2002). Fortunately, the majority of students have developed extensive computer skills. According to a NAEP survey, 85% of 4th grade students and 88% of 8th grade students use computers at home. In addition, 86% of 4th grade students and 80% of 8th grade students also use computers at school. These percentages are expected to rise in the future (NAEP 2005).

Summary

The research does not provide a clear answer with regards to whether aggregate level mode differences exist for constructed and selected response assessments administered via computer or paper-and-pencil. In general, there are no major effects of modality for either type of assessment. On an individual basis, differences in achievement appear to

exist due to differences in comfort using the computer and typing skills. The individual differences swing in both directions—some are advantaged by the computer while others are disadvantaged—primarily depending upon comfort with computers. Similarly, some students are advantaged by using paper and pencil, while others are disadvantaged with this mode of delivery.

Currently, most high-stakes assessments are still completed using paper-and-pencil. In the future, it is likely more and more assessments will be completed using the computer. This trend will likely parallel the increase in school computer use. The research currently supports this trend. Students who are more comfortable and familiar with computers are able to perform as well or better than they would have on a paper-based test. So, as computers become an increasingly integral part of the classroom, the research indicates that testing on the computer will become more appropriate than testing on paper. As is the case with the delivery of paper-and-pencil assessments, test publishers and administrators using computer-delivered tests must do everything possible to ensure that the assessment is fair for all students.

References

Bangert-Drowns, R.L. (1989). *Research on wordprocessing and writing instruction.* Retrieved July 12, 2006, from ERIC database.

Bridgeman, B., & Cooper, P. (1998). *Comparability of scores on word-processed and handwritten essays on the graduate management admissions test.* Retrieved July 12, 2006, from ERIC database.

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 2002 annual meeting of the National Council on Measurement in Education.*

Craig, D. (2001). *Handwriting legibility and word processing in assessing rater reliability.* Retrieved July 12, 2006, from http://www.danielcraig.com/illinois/mathesis/

Grejda, G.F., & Hannafin, M.J. (1992). Effects of Word Processing on Sixth Graders' Holistic Writing & Revisions. *Journal of Educational Research.* Retrieved July 12, 2006, from ERIC database.

Harrington, S., Shermis, M.D., & Rollins, A.L. (2000). The influence of word processing on English placement test results. *Computers and Composition.* Retrieved July 12, 2006, from ERIC database.

Manalo, J.R., & Wolfe, E.W. (2000). *The impact of composition medium on essay raters in foreign language testing.* Retrieved July 12, 2006, from ERIC database.

National Center for Education Statistics. (2005). *Online assessment in mathematics and writing: Reports from the NAEP technology-based assessment project, research and development series.* Retrieved July 12, 2006, from http://nces.ed.gov/nationsreportcard/pdf/studies/2005457.pdf

Nichols, L.M. (1996). Pencil and paper versus word processing: A comparative study of creative writing in the elementary school. *Journal of Research on Computing in Education.* Retrieved July 12, 2006, from ERIC database.

Parshall, C.G., & Kromrey, J.D. (1993). *Computer testing versus paper-and-pencil testing: An analysis of examinee characteristics associated with mode effect.* Retrieved July 12, 2006, from ERIC database.

Peek, P. (2005). *Recent trends in comparability studies.* Retrieved July 12, 2006, from http://www.pearsonedmeasurement.com/downloads/research/RR_05_05.pdf

Powers, D.E., Fowles, M.E., Farnum, M., & Ramsay, P. (1994). *Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays.* Retrieved July 12, 2006, from ERIC database.

Puhan, G., & Boughton, K.A. (2004). *Evaluating the comparability of paper and pencil versus computerized versions of a large-scale certification test.* Retrieved July 12, 2006, from ERIC database.

Russell, M., & Haney, W. (1996). *Testing writing on computers: Results of a pilot study to compare student writing test performance via computer or via paper-and-pencil.* Retrieved July 12, 2006, from ERIC database.

Russell, M., & Haney, W. (1997). *Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil.* Retrieved July 12, 2006, from ERIC database.

Russell, M., & Haney, W. (2000). The gap between testing and technology in schools. *The National Board on Education Testing and Public Policy.* Retrieved July 12, 2006, from ERIC database.

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.* Retrieved July 12, 2006, from ERIC database.

Wolfe, E.W., & Manalo, J.R. (2004). Composition medium comparability in a direct writing assessment of non-native English speakers. *Language Learning & Technology,* 81, 53-65.