



inTASC

TECHNOLOGY AND ASSESSMENT
STUDY COLLABORATIVE

Examining the Effect of Text Editor and Robust Word Processor on Student Writing Test Performance

**Part of the New England Compact
Enhanced Assessment Project**

Michael Russell, Jennifer Higgins, & Thomas Hoffmann
Technology and Assessment Study Collaborative
Boston College
332 Champion Hall
Chestnut Hill, MA 02467

www.intasc.org

Examining the Effect of Text Editor and Robust Word Processor Use on Student Writing Test Performance

Michael Russell, Jennifer Higgins, & Thomas Hoffmann
Technology and Assessment Study Collaborative
Boston College
Released November 2004

This study has been funded by the New England Compact Enhanced Assessment Project through US Department of Education Grant #S368A030014.

The New England Compact (Maine, New Hampshire, Rhode Island, Vermont) provides a forum for the states to explore ideas and state-specific strategies, build and expand a collective knowledge base, and initiate cross-state collaborative activities that benefit each state with economies of scale and cost-efficiency. A primary focus for the New England Compact is the implementation of the No Child Left Behind (NCLB) legislation. Current Compact projects include activities initiated by the Commissioners of Education, Deputy Commissioners, State Assessment Directors, and Title III and Special Education Directors and participation in an Enhanced Assessment Grant funded by the Federal Department of Education.
(www.necomact.org)

Copyright © 2004 Technology and Assessment Study Collaborative, Boston College



Examining the Effect of Text Editor and Robust Word Processor Use on Student Writing Test Performance

Michael Russell, Jennifer Higgins, and Thomas Hoffmann
Technology and Assessment Study Collaborative
Boston College
Released November 2004

Over the past decade increasing numbers of students have been using word processing throughout the writing process. During this same time, assessment and accountability programs have become an integral part of the education system through the No Child Left Behind legislation. These two combined influences of technology infusion and the rise of state testing programs create questions about the validity of writing assessments administered on pencil and paper for students who are accustomed to writing on computers and conversely the validity of writing assessments administered on computers for students who are accustomed to writing using pencil and paper. Several small scale studies suggest that students who write with word processors under-perform on writing tests that require handwritten responses (Russell & Haney, 1997; Russell, 1999; Russell & Plati, 2001). Additionally, students who are accustomed to writing on paper under-perform when they are required to write with word processors. Research also provides some evidence that this underperformance is greater for students whose individual education plans (IEPs) suggest that they compose text with word processors (Russell & Plati, 2000). Finally, this line of research has sparked questions about the effect that word processing tools, such as spell check and grammar check would have on student performance. The purpose of this study is to provide further evidence regarding the impact that the use of a word processor during assessment has on student writing.

The research described here details our effort to expand on the previous writing and computer research, using a larger sample size, across multiple states and diverse geographical settings. The study examines the effects of using paper and pencil for a writing assessment versus using two alternate text production tools: a) use of a word processor with automatic editing tools disabled; b) use of a word processor with automatic editing tools enabled.

Specifically, the research questions include:

1. When 8th grade writing assessments are administered on computers, what is the effect on students' test performance?
2. Does writing performance differ when students have access to editing tools such as spell check and grammar check?
3. What, if any, is the relationship between the modal effect and students' computer fluidity, literacy, and/or use?

This research was federally funded through the Enhancing State Assessment grant program and conducted collaboratively with the Rhode Island, Vermont, New Hampshire, and Maine Departments of Education, the Education Development Center (EDC), and CAST.

Background

As the use of computer-based technologies continues to become a regular component of classroom teaching and learning, the tools with which students solve problems and produce work are transitioning from pencil-and-paper-based to computer-based. Yet, as students become increasingly accustomed to learning and working with these computer-based tools, a misalignment is developing between the tools some students use to learn and the tools students are allowed to use while their achievement is tested.

By far, the most common educational use of computers by students is for writing. In 1998, a national survey of teachers found that 50% of K–12 teachers had students use word processors (Becker, 1999). Similarly, during the spring of 2002, a survey of over 4,000 K–12 teachers and 13,000 students in 144 Massachusetts schools found that 86% of teachers report that they have students use computers or portable word processors for writing in class at least one time during the year, 79% have students do projects or papers using a computer outside of class, and 81% of fifth-graders and 94% of middle and high school students report having produced first drafts on computer during the past year (Goldberg, O'Connor, & Russell, 2002).

Despite considerable debate as to the impact computer use has on student learning, there is a growing body of research that suggests that the impact is generally positive, particularly in the area of writing (Goldberg, Russell, & Cook, 2003). The research on computers and writing suggests many ways in which writing on computers may help students produce better work. Although much of this research was performed before large numbers of computers were present in schools, formal studies report that when students write on computer they tend to produce more text and make more revisions. Studies that compare student work produced on computer with work produced on paper find that for some groups of students writing on computer also has a positive effect on the quality of student writing. This positive effect is strongest for students with learning disabilities, early elementary students, and college students. Additionally, when applied

to meet curricular goals, education technology provides alternative approaches to sustaining student interest, developing student knowledge and skill, and provides supplementary materials that teachers can use to extend student learning. As one example, several studies have shown that writing with a computer can increase the amount of writing students perform, the extent to which students edit their writing (Dauite, 1986; Etchinson, 1989; Vacc, 1987), and, in turn, leads to higher quality writing (Hannafin & Dalton, 1987; Kerchner & Kisting, 1984; Williamson & Pence, 1989). More recently, a meta-analysis of studies conducted between 1990 and 2001 also shows that the quality of student writing improves when students use computers for writing over an extended period of time (Goldberg, Russell, & Cook, 2003).

Mode of Administration Effect and Computer-Savvy Students

Despite the increasing presence and use of computers in schools, particularly for writing, many state testing programs require that students produce responses to open-ended and essay questions using paper and pencil. 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, Massachusetts, 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 68 students taking math, science and language arts tests, including both multiple-choice and open-ended items, on paper, and another group of 46 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% performed at a “passing” level, but when they wrote on computer, 67% “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 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 200 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 4, 8 and 10. In addition, this most recent study also found that students accustomed to writing with eMates (portable writing devices capable of displaying about 20 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. Based on 1999 MCAS results, it was estimated that 19% of the fourth-graders classified as “Needs Improvement” would move up to the “Proficient” performance level. An additional 5% of students who were

classified as “Proficient” would be deemed “Advanced.” Similarly, it was estimated that in Grade 8, 4% of students would move from the “Needs Improvement” category to the “Proficient” category and that 13% more students would be deemed “Advanced.” Within one elementary school, it was estimated that the percentage of students performing at or above the “Proficient” level would nearly double from 39% to 67%.

Despite this body of evidence that consistently indicates that paper-based writing tests underestimate the writing performance of students who are accustomed to writing with computers, questions remain about the generalizability of these findings. Given that this body of research was conducted in only eight schools located in two Massachusetts districts, it is important to examine the extent to which these effects might occur in a larger and more diverse sample of students. In response to this need, the study presented here examines the mode of administration effect using a sample of approximately 1,000 eighth grade students who attend schools in rural, urban and suburban areas in Rhode Island, Vermont, and New Hampshire.

Design

To examine whether and how the use of simple text editors or robust word processors (that include spell and grammar checks) affect the writing test performance of 8th grade students, 966 students from twelve schools located in New Hampshire, Vermont, and Rhode Island were randomly assigned to perform the same writing test in one of three modes: 1) on paper, 2) on computer with a basic text editor; and 3) on computer with Microsoft Word with spell check and grammar check enabled. The participating schools were selected with the cooperation of the state Directors of Assessment. When selecting schools, we aimed to acquire a mix of rural, suburban, and urban schools and a mix of students that included English Language Learners (ELL), students identified with special needs in the area of language, as well as non-special education native English speaking students. Since ELL students tended to be clustered within schools, and since the location of the school could not be manipulated, random assignment occurred within rather than across schools. Through this three group randomized design, the research compares the effect of using paper and pencil versus two types of word processors on student test performance.

To control for effects that might result from differences in the type of word processors available within each school, the research team brought into each school a set of Macintosh laptop computers (G3 iBooks, 12-inch screens) with traditional hand-held mice that ran Microsoft Word. All students performing the writing test using the robust word processor used one of the research team’s laptop computers. Students assigned to the basic word processor group used a school computer, but were required to compose their response in a web-based text editor that was developed specifically for this study by the research team. In addition to performing

the same extended writing task, all students completed a computer fluidity test, a computer literacy test, and a computer use survey. The computer fluidity test was administered to all students on a laptop or over the Internet using the school's computers. The computer literacy and the computer use surveys were administered to all students on paper. The purpose of administering these three additional instruments was to collect multiple measures of students' computer skills, knowledge, and use so that we could examine the extent to which any modal affects are related to differences in students' ability or familiarity with using a computer – constructs that are not intended to be measured by the writing test.

Data Collection

Data were collected between January 24th and February, 28th, 2004. Within each school, students were randomly assigned to one of three test conditions. Each test condition was administered in a separate room (or set of rooms). Students assigned to compose their essay on paper were placed in regular classrooms and worked at a typical desk. Students in the paper group performed the computer fluidity test using laptop computers connected to a mouse, which were provided by the research team. Students assigned to compose their essay using the basic word processor were assigned to the school's computer laboratory and composed their essay using the school's desktop computers. To ensure that students assigned to the robust word processor group had access to the same set of word processing tools, this group of students composed their essays using laptop computers (i.e., G3 iBooks with 12-inch screens running Microsoft Word). For each test condition, as students entered the room, they were asked to find their place by looking for their name card on a desk or computer, which was set up with either the paper and pencil assessment or with the launched assessment application on the computer. Researchers then explained the purpose of the research and briefly described the writing assessment, fluidity exercises, computer literacy test, and survey to the students. Students were given one hour to complete the writing assessment and an additional hour to complete the computer fluidity, literacy, and use instruments.

Valid data was collected from 966 8th Grade students. The number of students participating from each school ranged from 25 to 188. Table 1 summarizes demographic information about students participating in the study. Specifically, Table 1 indicates that 110 students (11.4%) attended rural schools, 608 (63.1%) attended suburban schools, and 245 (25.4%) attended urban schools. In addition, 11.1% of students reported that they spoke a language other than English at home and 13.2% of students indicated that they had an IEP.

(Table 1 is shown on the following page.)

Table 1: Demographic Summary for Participating Students

Geography	Percent of Sample (n)
Rural	25.4 (110)
Suburban	63.1% (608)
Urban	11.4% (245)
Gender*	
Male	51.2
Female	48.8
Ethnicity/Race*	
White	81.1
African-American	2.9
Hispanic or Latino	3.1
Asian or Pacific Islander	2.3
Native American	1.3
Other	9.3
Multilingual Status*	
Multilingual	11.1
Nonmultilingual	88.9
IEP*	
Yes	13.2
No	27.0
Do Not Know	59.8

* As reported by participating students.

Instruments

Students participating in this study completed four data collection instruments in the following order: 1) Writing Test; 2) Computer Fluidity Test; 3) Computer Literacy Test; and 4) Computer Use Survey. Below, each instrument is described in greater detail. In addition, we explain in detail how scale scores were developed for the fluidity, literacy, and use instruments and present information on the reliability and validity of each scale.

Writing Test

The writing test consisted of a single essay prompt that had been used previously as part of the Maine Grade 8 Language Arts Test. The item asked students to respond to the following prompt: *Write about an important lesson that children should learn.* In addition to the prompt, students were provided with a sheet of paper that told them that their essay would be scored by two readers and that each reader would assign two separate scores. One score would focus on stylistic and rhetorical aspects of their writing and one score would focus on standard English conventions. Students were then presented with additional scoring information that provided more details on the criteria the readers would consider when awarding scores for the stylistic qualities of their writing and use of standard English conventions. Students were told that they had an hour to write as much as they could in response to this prompt and that they could write a first and second draft, if desired.

When scoring the student responses, the same scoring criteria used by the state of Maine were employed for this study. Idea/Topic Development was scored on a six-point scale, with 6 representing the highest score. Standard English Conventions was scored on a four point scale.

Eight readers, all of whom are or recently have been teachers in middle and/or high school settings, were employed for the study. All readers were also blind to the study's purpose. Before any essays were scored, all readers participated in the same four hour training session during which the scoring guide was introduced and readers' scored approximately 50 practice essays. After a set of 10 essays was scored, the scores awarded by all readers were compared and essays that resulted in more than a one point discrepancy were discussed at length in order to help all readers interpret and apply the scoring criteria in the same manner. This process was repeated until a set of 10 essays was scored and less than 10% of scores differed by more than one point. After the training was complete, readers were paired, but were not informed of their partner, and were asked to score approximately 275 essays per pair. Each pair of readers scored an equal number of essays that were originally composed on paper, using a text editor, or Microsoft Word. Essays composed in each medium were intermixed so that effects resulting from fatigue or changes in severity or leniency that occur over time would not be correlated with the essay medium. As seen in Table 2, the correlation between scores awarded

by two raters was .82 for the Topic/Idea Development scale and .67 for the Standard English Conventions scale. In addition, 99.2% of essays were awarded scores within 1 point for the Topic/Idea Development scale and 99.8% of essays were awarded scores within one point for the Standards English Convention scale. Although Kappa was below .5, it is important to note that the scale for Topic/Idea Development contained 6 score points. Thus a Kappa of .40 is considerably higher than the .167 probability of two raters awarding the same score by chance alone. Of some concern, though is the relatively low Kappa for the Standards English Conventions scale. While Standard English Conventions is scored on a four-point scale and a Kappa of .37 is statistically significantly higher than the .25 probability of two raters awarding the same score by chance, the relatively small difference between .25 and .37 suggests that agreement between raters for this scale was lower than is typically desired.

Table 2: Inter-Rater Reliability

	Correlation (r)	% Exact Match	% Within 1 point	Kappa
Topic/Idea Development	.82	54.9	99.2	.40
Standard English Conventions	.67	58.7	99.8	.37

Description of Data Collection Conditions and Instruments

When composing their essays, students within each participating school were randomly assigned to one of three groups. Students assigned to the paper group composed their essay on lined paper. Students were provided with six sheets of lined paper and were told that they could request more pages if needed. Students assigned to the Text Editor group composed their essay directly on a computer using a basic text editor. The text editor allowed students to copy, cut, and paste text, but did not allow them to format text or to access a spell-check or grammar-check. The text editor program ran on the Internet and students writing was saved directly to a server connected to the Internet. Students assigned to the Word Processor group composed their essays using Microsoft Word on 12-inch iBook laptop computers. Students were provided full access to spell-check and grammar-check and were allowed to change fonts and font sizes if desired.

After all data collection was complete, essays composed on paper were transcribed verbatim (including spelling, capitalization, and punctuation errors) into computer text. In addition, all formatting and font changes made by students in the word processing group were removed. All essays were then formatted in the same way using 12 point Times Roman single-spaced text and were printed on paper for readers to score.

Computer Fluidity Test

After completing the writing assessment, all students were asked to complete a computer fluidity test that consisted of four sets of exercises. The purpose of the computer fluidity test was to measure students' ability to use the mouse and keyboard to perform operations similar to those they might perform on a test administered on a computer. In this report, we refer to these basic mouse and keyboard manipulation skills as "computer fluidity".

The first exercise focused on students' keyboarding skills. For this exercise, students were allowed two minutes to keyboard words from a given passage. The passage was presented on the left side of the screen and students were required to type the passage into a blank text box located on the right side of the screen. The total number of characters that the student typed in the two-minute time frame was recorded. A few students typed words other than those from the given passage. These students' were excluded from analyses that focused on the relationship between writing test performance and computer fluidity.

After completing the keyboarding exercise, students performed a set of three items designed to measure students' ability to use the mouse to click on a specific object. For these items, students were asked to click on hot air balloons that were moving across the computer screen. The first item contained two large hot air balloons. The second item contained five medium-sized hot air balloons that were moving at a faster rate. The third item contained 10 small hot air balloons that were moving at an even faster rate. In each set of hot air balloons, the amount of time and the number of times the mouse button was clicked while clearing the screen were recorded.

The third computer fluidity exercise measured students' ability to use the mouse to move objects on the screen. For this exercise, students were presented with three items each of which asked students to drag objects from the left hand side of the screen to a target on the right hand side of the screen. For the first item, students were asked to drag books into a backpack. The second item asked students to drag birds into a nest. The third item asked students to drag ladybugs onto a leaf. As students advanced through the drag and drop levels, the size of both the objects and the targets decreased, making the tasks progressively more difficult. Similar to the clicking exercise, for each item the amount of time and the number of times the mouse was clicked were recorded.

Finally, the fourth exercise was designed to measure how well students were able to use the keyboard's arrow keys to navigate on the screen. For this exercise, students were asked to move a ball through a maze by using the arrow keys. Students were shown where on the keyboard to find the arrow keys. The first half of the maze consisted of 90-degree turns and the second half contained turns with curves. The time required to reach the end of the maze were recorded. As described later in this report, the data from the keyboarding, clicking, drag and drop, and arrow key exercises was combined into a single scale to produce a computer fluidity score for each student.

Computer Literacy Test

After finishing the computer fluidity exercises, students were asked to complete a short paper-based computer literacy test. The purpose of this test was to measure students' familiarity with computing terms and functionality. Virginia and North Carolina have administered multiple choice computer literacy tests to students at the eighth-grade level. Fifteen released multiple-choice items from previously administered Virginia and North Carolina assessments were used in the computer literacy test as part of this research. Items were chosen based on their alignment with the International Society for Technology in Education standards.

Computer Use Survey

Lastly, students were asked to complete a paper-based survey. This survey was adapted from the grade five student survey constructed for the Use, Support, and Evaluation of Instructional Technology (USEIT) study (see Russell, Bebell and O'Dwyer, 2003). Students were asked questions focusing on their specific uses of technology in school and at home, their comfort level with technology, as well as some demographic information.

Scale Development

Three instruments were administered to students in order to measure their computer fluidity, computer literacy, and computer use. Each of these instruments was developed specifically for this study. While items that comprised the literacy and use instruments were taken directly from instruments that have been used in previous research and/or state test administrations, the specific set of items that comprise each instrument had not been combined and used in this form. In addition, the items that formed the computer fluidity test were developed by the research team and had not previously been administered to a large number of students. Thus, before information from these three instruments could be used for analytic purposes, scale scores had to be developed and the reliability of these scales was examined. To this end, two sets of analyses were conducted to create and then examine the reliability of these scales. First, principal component analyses were performed on each instrument in order to examine the extent to which the items could be grouped to form a single score. In cases where all items could not be combined to form a single scale, principal component analyses were used to identify a subset of items that formed a unidimensional scale. Scale scores were then created for each student. Second, Cronbach's alpha was calculated for each scale to examine the reliability of the scale. In cases where the scale had unacceptably low reliability (below .60), item to total score correlations were examined to identify items that were contributing to low reliability. These items were then dropped from the scale, new scale scores were created, and the reliability analysis was repeated. Below, we describe the specifics for each scale creation.

Computer Fluidity Scale

The computer fluidity test consisted of five sets of tasks. As described in the instrument section, the five tasks included: keyboarding, clicking, drag and drop, navigating with the arrow keys, and completing a cut and paste exercise. The keyboarding, arrow key, and cut and paste tasks consisted of a single item and the only data recorded pertained to the number of characters typed for the keyboarding and the amount of time required to complete the task for the arrow key and cut and paste. The two other tasks each consisted of three items. For each item, two pieces of information were collected: a) the amount of time required to complete the item, and b) the number of mouse clicks required to complete the item. First, all fluidity components were standardized to create variables that were on the same scale. Then all standardized variables were employed to compute a scale score. However, through an initial principal component analyses, it was clear that some of this information was not useful. To maximize the variance retained by combining information from each of the computer fluidity items into a single score, only the time information from each task was used. When a factor analysis of the balloon time variables, the drag times, the arrow time, the reversed keyboarding time, and the cut and paste time was conducted, a one factor solution that accounted for 39.7% of the total variance with an alpha reliability of 0.80 was achieved. Factor loadings of each variable used range from 0.48 – 0.77.

It is noteworthy that in one administration, network problems prevented accurate data collection of the fluidity data. The data for these students, as well as the data for students who keyboarded words other than those from the given passage was not used in the fluidity factor and reliability analyses.

Computer Literacy Scale

The computer literacy test consisted of 15 multiple-choice items that asked students about specific aspects of a computer. These aspects included terminology, software, hardware, and tasks typically performed with a computer. When a principal components analysis was run on the 15 items, a three-factor solution emerged. One item was removed due to its low point biserial correlation. A principal components analysis was then run on the 14 remaining items and one dominant factor emerged. The one-factor solution accounted for 21.0% of the variance and had an alpha reliability coefficient of 0.70. Factor loadings on the 14 items ranged from 0.35 – 0.64. This one factor solution was used to create scaled scores of students' computer literacy.

Home Computer Use Scale

To measure the extent to which students used a computer at home, a series of questions on the student computer use survey asked how frequently they use computers at home to play games, chat/instant message, email, search the Internet for school, search the Internet for fun, listen to mp3/music, write papers for school, and/or create/edit digital photos or movies. Students were asked to choose one of

the following responses for each activity: never, about once a month, about once a week, a couple of times a week, and every day.

When a principal components factor analysis was run on the eight home computer use items, a two-factor solution emerged. The first factor represented fun computer uses at home (play games, chat/instant message, email, search the Internet for fun, listen to mp3/music, and create/edit digital photos or movies). The second factor represented academic computer uses at home (search Internet for school and write papers for school). A principal components factor analysis was then run on the six fun home computer use items, yielding a one factor solution that accounted for 53.4% of the variance and had an alpha reliability of 0.82. Factor loadings on the six items ranged from 0.61 – 0.80. This one factor solution was used to create scaled scores of students' home computer use for fun.

School Computer Use Scale

To measure the extent to which students use computers in school, a series of questions on the student computer use survey asked how frequently they use computers in school to email, write first drafts, edit papers, find information on the Internet, create a Hyperstudio or Powerpoint presentation, play games, and/or solve problems. Students were also asked how frequently they use a computer in school for reading/language arts, social studies, math, and science. For both sets of items, students were asked to choose one of the following responses for each activity: never, about once a month, about once a week, a couple of times a week, and every day. When a principal components factor analysis was run on the 11 school computer use items, a three-factor solution emerged. One item (problem solving) was removed due to low point biserial correlation. When a principal components factor analysis was run on the 10 remaining school computer use items, a two-factor solution emerged. The first factor represented academic computer uses at school (write first drafts, edit papers using a computer, find information on the Internet, create Hyperstudio or Powerpoint presentations, reading/language arts, social studies, math, and science). The second factor represented fun computer uses at school (email someone and play games). A principal components factor analysis was then run on the eight academic school use items, yielding a one factor solution that accounted for 40.5% of the variance and had an alpha reliability of 0.78. Factor loadings on the eight items ranged from 0.46 – 0.74. This one factor solution was used to create scaled scores of students' school computer use for academics.

Writing Scale

To measure the extent to which students use computers for writing, students were asked how frequently they use a computer to write first drafts in school and edit papers using a computer in school. Students were asked to choose one of the following responses for each activity: never, about once a month, about once a week, a couple of times a week, and every day. When a principal components factor analysis was run on the two writing computer use items, a one-factor solution emerged which accounted for 76.2% of the variance and had an alpha reliability of 0.68. The factor loading on each of the two items was 0.68. This one factor solution was used to create scale scores of students' use of computers for writing.

Results

To examine the extent to which the mode of testing writing skills affect student performance, a series of analyses were performed. These analyses include a comparison of mean Topic Development, Standard English Conventions, and total writing scores across the three modes. To examine the extent to which prior experience and skill using a computer interacted with the composition mode, analyses were conducted using the keyboarding, computer fluidity, literacy, and use measures. For each of these measures, students were divided into three groups representing high, medium, and low levels of keyboarding, fluidity, literacy, or use. The modal effect was then examined within each group by comparing performance across the three presentation modes. In addition, regression analyses were conducted to examine the modal effect after adjusting for differences in prior fluidity, literacy, and use. Finally, to examine the extent to which the modal effect differed by demographic variables, a series of analyses was conducted that focused on gender, race/ethnicity, language, and special education. Findings from each of these analyses are presented below. Note that in the analyses presented below, students who composed an essay on paper are referred to as the Paper group. Students who composed essays on a computer using a basic text editor that did not allow changes in fonts or access to spell-check or grammar-check are referred to as Basic WP. Students who composed essays using Microsoft Word with full access to spell-check and grammar-check are referred to as Full WP (i.e., full-featured word processor).

Writing Scores

As seen in Table 3, the mean topic development score was 6.03 on a scale that ranged from 2 to 12. The mean English standards score was 5.02 on a scale that ranged from 2 to 8. The mean total score (which is the combination of topic development and English standards scores) was 11.06 on a scale that ranges from 4 to 20.

Table 3: Comparison of Scores Among the Three Composition Modes for All Students

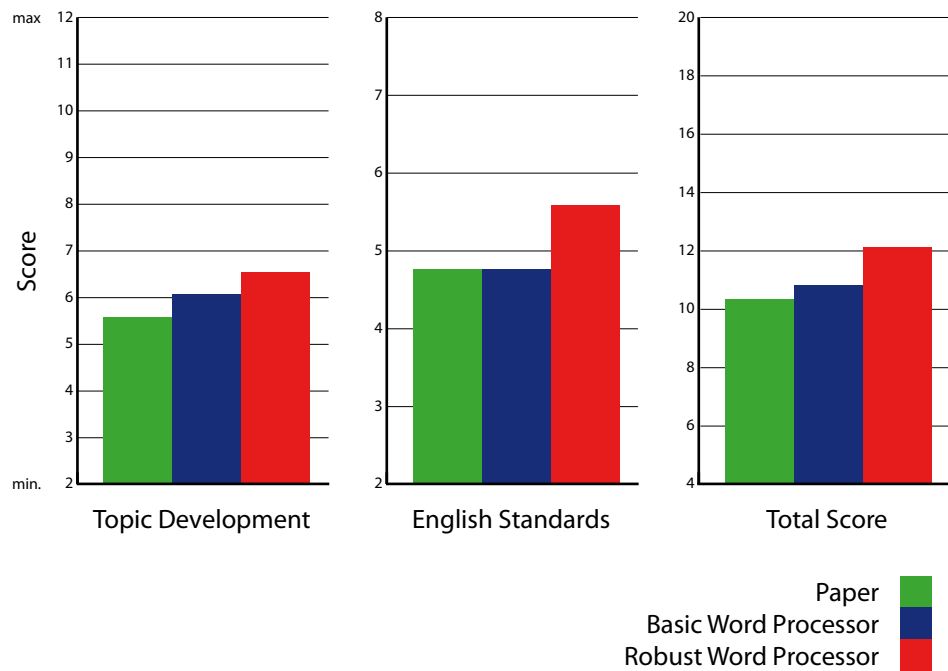
Group	N	Mean	St. Dev.	F	df (within, between)	p
Topic Development						
All Students	966	6.03	2.10			
Paper ^{1,2}	344	5.56	2.03	18.27	963, 2	<.001
Basic WP ³	312	6.06	1.97			
Robust WP	310	6.53	2.19			
English Standards						
All Students	966	5.02	1.45	36.73	963, 2	<.001
Paper ²	344	4.76	1.32			
Basic WP ³	312	4.76	1.42			
Robust WP	310	5.58	1.45			
Total Score						
All Students	966	11.06	3.31	26.68	963, 2	<.001
Paper ²	344	10.32	3.07			
Basic WP ³	312	10.82	3.16			
Robust WP	310	12.12	3.43			

1 significant difference between Paper and Basic WP at $p < .05$ using Tukey's post-hoc mean comparison

2 significant difference between Paper and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

3 significant difference between Basic WP and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

As seen in Table 3, there is a significant difference in scores among groups for topic development, English standards, and total writing scores. As depicted in Figure 1, students who composed their essay using a robust word processor received significantly higher scores than did students who used paper or a basic word processor. Students who composed their essay using a basic word processor as compared to using paper also received significantly higher scores for the Topic Development dimension, but not for the English Standards dimension or for the total score. Based on a total group analysis, it appears that students produce essays that have better developed topics when they are able to use a text editor than when they use paper. In addition, students who have access to full-featured word processing tools produce essays of higher quality than do students who use a text editor without access to a full-range of word processing tools.

Figure 1: Mean Scores by Testing Mode

Examining Score Differences by Prior Computer Skills and Use

Recognizing that students' computer skills and prior experiences using a computer for writing may affect their performance when composing text on computer or on paper, a series of analyses was conducted to examine the relationship between students' performance on the writing test with their performance on the keyboarding test, fluidity test, computer literacy test, and computer use survey. If students who have strong computer skills, knowledge, and use tend to perform better when they compose written responses on a computer and students with weaker computer skills and fewer prior experiences working with computers perform worse when composing essays on computer, then students' computer experiences may interact with the mode of composition. In order to examine this potential relationship, students were divided into three groups based on their scale scores. This grouping was conducted separately for the keyboarding, fluidity, literacy, home use, and school use scales. For each scale, the three groups represented high, medium, and low level scores. For each scale, students in the top third of the distribution were placed into the high group. Students whose score fell in the middle third of the distribution were placed into the medium group. Students whose scale score was in the lowest third of the distribution were placed into the low group. After these three groups were formed, the three writing scores were compared across the three composition modes based on the high-medium-low classification of each computer measure.

Although students were randomly assigned to the three mode of composition groups, differences in students' prior computer experiences may still exist. If so, these differences may mask differences that result due to the presentation mode. To examine differences in prior computer experience, mean scores for the keyboard, fluidity, literacy, home use, and school use scales were compared among the three mode of presentation groups. As seen in Table 4 and depicted in Figure 2, keyboarding scores were significantly lower for the paper group as compared to the Basic WP and Robust WP groups. Similarly, computer fluidity scores were significantly lower for the paper group as compared to the two other groups. In addition, computer literacy scores were significantly higher for the Basic WP group than the two other groups. For the computer literacy scores, the paper group was significantly lower than the Robust WP group and was lower than the Basic WP group (although this score difference was not statistically significant). There were no statistically significant differences among groups for any of the computer use scales.

Table 4: Comparison of Scale Scores Across Presentation Modes

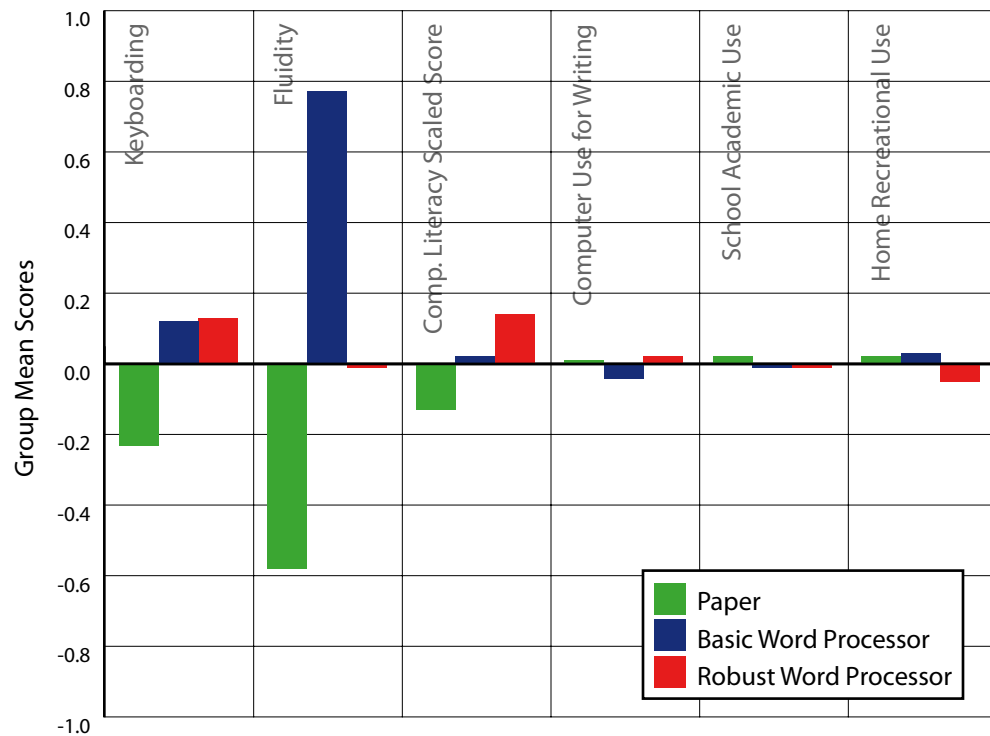
Scale	Paper Mean	Basic WP Mean	Robust WP Mean	Pooled SD	F	Sig.
Keyboarding ^{1,2}	-.23	.12	.13	1.00	13.05	<.001
Fluidity ^{1,2,3}	-.58	.77	-.01	1.00	167.06	<.001
Comp Literacy Scaled Score ²	-.13	.02	.14	1.00	5.10	.006
Computer Use for Writing	.01	-.04	.02	1.00	.17	.84
School Academic Use	.02	-.01	-.01	1.00	.06	.95
Home Recreational Use	.02	.03	-.05	1.00	.49	.62

1 significant difference between Paper and Basic WP at $p < .05$ using Tukey's post-hoc mean comparison

2 significant difference between Paper and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

3 significant difference between Basic WP and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

Figure 2: Group Mean Scores for Keyboarding, Computer Fluidity, Computer Literacy, and Computer Use Measures



To control for differences among the three groups for each of these measures, regression analyses were performed. For each regression analysis, the computer skill or use measure was entered as an independent variable, and two dummy variables were entered to estimate the effect of group membership on test scores controlling for differences in computer skill or use score. Specifically, the first dummy variable represented the Basic WP group (i.e., students in the Basic WP group were assigned a 1 and students in the two other groups were assigned a 0). The second dummy variable represented the Robust WP group (i.e., students in the Robust WP group were assigned a 1 and students in the two other groups were assigned a 0). Students' writing test scores were the dependent variables in all the regression analyses.

The full set of analyses for the keyboarding, computer fluidity, computer literacy, home use, and school use scales are presented separately below.

Keyboarding Speed Analysis

As described above, students were assigned to a high, medium, or low keyboarding speed group based on their keyboarding score. Table 5 displays the average writing test score for students based on their keyboarding categorization, across each of the three modes. Note that for these analyses, only students who have both valid writing scores and a valid keyboarding score are included.

Table 5: Comparison of Mean Writing Test Score across Keyboarding Speed Groups and Presentation Mode

	N	All Students	Paper	Basic WP	Robust WP	F	P
Topic Dev.							
Low	296	4.95	4.89	4.88	5.14	.51	.60
Medium ²	296	6.26	5.94	6.17	6.68	4.25	.02
High ^{1,2}	291	7.08	6.28	7.22	7.53	9.85	<.001
Eng. Stand.							
Low ^{1,3}	296	4.30	4.43	3.81	4.72	11.97	<.001
Medium ^{2,3}	296	5.21	5.02	4.98	5.64	8.22	<.001
High ^{2,3}	291	5.62	5.05	5.46	6.19	17.31	<.001
Total Score							
Low ³	296	9.25	9.32	8.68	9.86	3.63	.03
Medium ^{2,3}	296	11.47	10.96	11.15	12.32	6.42	.002
High ^{1,2,3}	291	12.7	11.33	12.68	13.71	13.87	<.001

1 significant difference between Paper and Basic WP at $p < .05$ using Tukey's post-hoc mean comparison

2 significant difference between Paper and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

3 significant difference between Basic WP and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

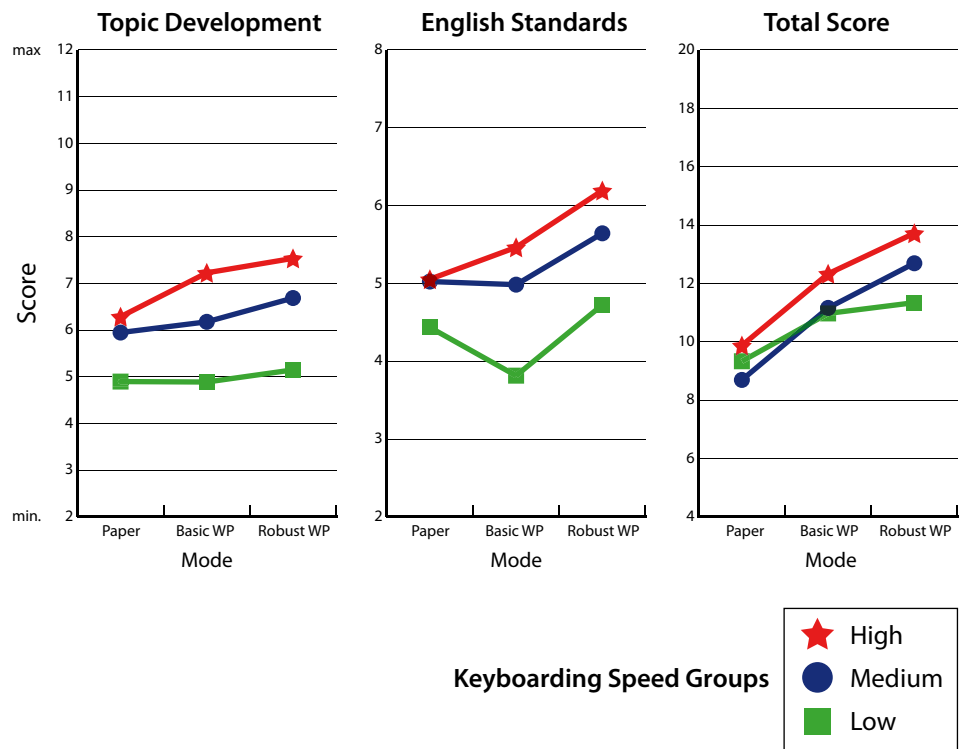
In general, students with higher keyboarding scores tended to score higher on the writing assessment across all three modes. Similarly, students in the low keyboarding group performed lower, on average, on each of the presentation modes. Since this relationship holds across all three testing modes, it does not provide evidence that students taking the assessment using the two computer-based modes were advantaged or disadvantaged by their keyboarding skills.

To examine differences within each keyboarding group across presentation modes, a one way ANOVA was run across presentation modes for each keyboarding categorization. For the topic development scores, there were no significant differences found across modes for the low category, but there were significant differences found at the medium ($p = .02$) and high ($p < .001$) categories. At the medium category, the differences occurred between the paper and robust word processing group, with the robust word processing group showing the higher average scores. While at the high category, there were differences between paper and both of the two computer modes. Again, the computer groups received higher scores than the paper group.

For the English standards scores, significant differences were found across all three categories at the $p < .001$ level. For the low category, the differences occurred between the paper and basic word processing group and between the basic word processing and robust word processing groups, with the basic word processing group achieving the lower scores in both comparisons. For the medium and high categories, there were differences between paper and robust word processing and between basic word processing and robust word processing, with the robust word processing group achieving higher scores.

For the total writing score, there were significant differences found at the low ($p = .03$), medium ($p = .002$), and high ($p < .001$) categories. At the low category, the differences occurred between the basic word processing and robust word processing groups, with the robust word processing group achieving the higher average score. At the medium category, there were differences between paper and robust word processing and between basic word processing and robust word processing, with the robust word processing group showing higher scores. At the high category, there were significant differences across all three groups, with the paper group achieving the lowest average score and the robust word processing group achieving the highest average score.

Figure 3: Mean Writing Scores for Keyboarding Groups and Presentation Mode



As described above, a regression analysis was performed in order to control for differences in keyboard ability that existed among the presentation modes. In addition to keyboard score, the regression model included two dummy variables which represented the basic and robust word processing groups. For the first regression model students' topic development score was the dependent variable. As seen in Table 6, the regression model accounted for 20% of the variance in topic development scores. As indicated by the standardized coefficient (Beta), keyboard ability was the strongest predictor of students' writing scores. Both basic and robust word processing conditions had a positive effect on the topic development score, after controlling for differences in keyboarding ability. However, only the coefficient for the robust word processing was statistically significant ($p < .001$).

For the second regression model students' English conventions score was the dependent variable. As seen in Table 6, the regression model accounted for 21% of the variance in standard English convention scores. As indicated by the standardized coefficient (Beta), keyboard ability was the strongest predictor of students' writing scores. The basic word processing condition had a negative effect on the standard English convention score, while the robust word processing condition had a positive effect the standard English conventions score, after controlling for differences in keyboarding ability. Both the basic word processing coefficient ($p = .03$) and the robust word processing coefficient ($p < .001$) were found to be statistically significant.

Finally in the third regression model students' total essay score was the dependent variable. As seen in Table 6, the regression model accounted for 23% of the variance in writing scores. As indicated by the standardized coefficient (Beta), keyboard ability was the strongest predictor of students' writing scores. Both basic and robust word processing conditions had a positive effect on the topic development score, after controlling for differences in keyboarding ability. However, only the coefficient for robust word processing was statistically significant ($p < .001$).

(Table 6 is shown on the following page.)

Table 6: Regression Model for Keyboarding Speed and Group Membership Predicting Reading Scores

Topic Development					
R ² = .20 F = 74.89 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	5.82	.11		53.28	<.001
Keyboarding	.88	.06	.42	13.67	<.001
Basic WP	.19	.16	.04	1.22	.22
Robust WP	.64	.16	.14	4.04	<.001
English Standards					
R ² = .21 F = 78.90 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	4.93	.08		65.41	<.001
Keyboarding	.56	.04	.39	12.69	<.001
Basic WP	-.23	.11	-.08	2.16	.03
Robust WP	.59	.11	.19	5.45	<.001
Total Score					
R ² = .23 F = 88.01 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	10.75	.17		63.30	<.001
Keyboarding	1.44	.10	.43	14.43	<.001
Basic WP	-.04	.24	.00	.17	.86
Robust WP	1.23	.25	.17	5.02	<.001

Although students were grouped by keyboarding speed and ANOVAs were performed within each keyboarding level to examine differences among composition modes, there remained a statistically significant difference in keyboarding speed among composition modes for the high keyboarding group. As seen in Table 7, students in the high keyboard group who composed their essay on paper had lower keyboarding speed than did students in the two computer groups. For this reason, the regression analyses presented above were performed within the high keyboarding group to examine the modal effects while controlling for differences in keyboard speed.

Table 7: Comparison of Mean Keyboard Speed Among Testing Modes Within Each Keyboard Level

	N	All Students	Paper	Basic WP	Robust WP	F	p
Low	296	-.97	-1.01	-.93	-.97	.95	.39
Medium	296	-.13	-.15	-.09	-.13	2.43	.09
High ^{1,2}	291	1.12	.83	1.27	1.18	8.13	<.001

1 significant difference between Paper and Basic WP at $p < .05$ using Tukey's post-hoc mean comparison

2 significant difference between Paper and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

3 significant difference between Basic WP and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

As seen in Table 8, the first regression model for the high keyboarding group uses students' topic development score as the dependent variable. The regression model accounted for 7% of the variance in topic development scores. As indicated by the standardized coefficient (Beta), the robust word processing condition was the strongest predictor of students' topic development scores. Both basic and robust word processing conditions had a positive effect on the topic development score, after controlling for differences in keyboarding ability. The basic word processing coefficient ($p = .01$) and the robust word processing coefficient ($p < .001$) were both found to be statistically significant.

For the second regression model of the high keyboarding group students' standard English conventions score was the dependent variable. As seen in Table 8, the regression model accounted for 14% of the variance in standard English convention scores. As indicated by the standardized coefficient (Beta), the robust word processing condition was the strongest predictor of students' writing scores. Both basic and robust word processing conditions had a positive effect on the English conventions score, after controlling for differences in keyboarding ability. However only the robust word processing coefficient ($p < .001$) was found to be statistically significant.

Finally in the third regression model for the high keyboarding group, students' total essay score was the dependent variable. As seen in Table 8, the regression model accounted for 10% of the variance in writing scores. As indicated by the standardized coefficient (Beta), the robust word processing condition was

the strongest predictor of students' writing scores. Both basic and robust word processing conditions had a positive effect on the topic development score, after controlling for differences in keyboarding ability. The basic word processing coefficient ($p=.02$) and the robust word processing coefficient ($p<.001$) were both found to be statistically significant.

Table 8: Regression Model for Keyboarding Speed and Group Membership Predicting Writing Scores for Students in the High Keyboarding Group

Topic Development					
$R^2 = .07$		$F = 7.95$		$P < .001$	
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	6.04	.25		24.16	<.001
Keyboarding	.30	.15	.12	1.99	.05
Basic WP	.81	.30	.20	2.73	.01
Robust WP	1.14	.29	.28	3.94	<.001
English Standards					
$R^2 = .14$		$F = 16.15$		$P < .001$	
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	4.75	.17		27.72	<.001
Keyboarding	.36	.10	.20	3.53	<.001
Basic WP	.25	.20	.08	1.22	.23
Robust WP	1.01	.20	.35	5.07	<.001
Total Score					
$R^2 = .10$		$F = 12.12$		$P < .001$	
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	10.79	.39		27.59	<.001
Keyboarding	.66	.23	.16	2.82	.005
Basic WP	1.05	.46	.16	2.28	.02
Robust WP	2.15	.45	.33	4.74	<.001

Computer Fluidity Analysis

As described above, students were assigned to a low, medium, or high computer fluidity group based on their fluidity scaled score. Table 9 displays the average topic development, English conventions, and total writing score for students based on their computer fluidity categorization, across each of the three modes.

Table 9: Comparison of Mean Writing Test Score across Computer Fluidity Groups and Presentation Mode

	N	All Students	Paper	Basic WP	Robust WP	F	p
Topic Development							
Low	273	5.48	5.36	5.47	5.73	0.93	.40
Medium ²	273	6.22	5.91	5.80	6.62	4.14	.02
High ^{2,3}	273	6.84	6.27	6.67	7.39	5.26	.006
English Standards							
Low	273	4.79	4.73	4.59	4.99	1.1.14	.32
Medium ^{2,3}	273	5.15	4.86	4.50	5.59	13.83	<.001
High ^{2,3}	273	5.38	5.05	5.05	6.24	21.69	<.001
Total Score							
Low	273	10.27	10.09	10.06	10.72	1.13	.33
Medium ^{2,3}	273	11.37	10.76	10.30	12.21	8.28	<.001
High ^{2,3}	273	12.22	11.32	11.72	13.63	12.50	<.001

1 significant difference between Paper and Basic WP at $p < .05$ using Tukey's post-hoc mean comparison

2 significant difference between Paper and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

3 significant difference between Basic WP and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

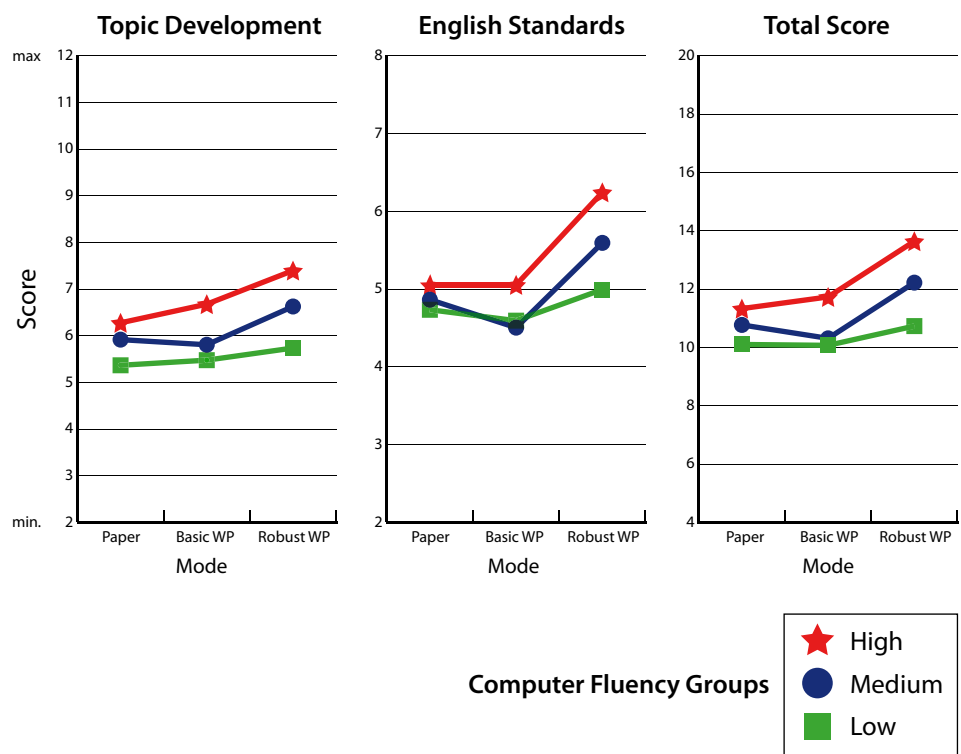
In general, students with higher computer fluidity scores tended to score higher on the writing assessment across all three modes. Similarly, students in the low computer fluidity group performed lower, on average, across each of the presentation modes. Since this relationship holds across all three testing modes, it does not provide evidence that students taking the assessment using the two computer-based modes were advantaged or disadvantaged by their computer fluidity.

To examine differences within each computer fluidity group across presentation modes, a one-way ANOVA was run across presentation modes for each fluidity categorization. For the topic development scores, there were no significant differences in scores at the low level, but there were statistically significant differences in scores at the medium ($p = .02$) and high ($p = .006$) levels. At the medium level, the differences were found between the paper group and the robust word processing group. At the high level, the differences were found between paper and robust word processing and between the basic word processing and robust word processing. In each case, the robust word processing group achieved the higher average score.

For the English conventions scores, there were no significant differences in scores at the low level, but there were statistically significant differences in scores at the medium ($p < .001$) and high ($p < .001$) levels. At both the medium and high levels, the differences were found between paper and robust word processing and between the basic word processing and robust word processing. In each case, the robust word processing group achieved the higher average score.

Similarly for the total writing scores, there were no significant differences in scores at the low level, but there were statistically significant differences in scores at the medium ($p < .001$) and high ($p < .001$) levels. At both the medium and high levels, differences were found between paper and robust word processing and between the basic word processing and robust word processing. In each case, the robust word processing group achieved the higher average score.

Figure 4: Mean Writing Scores for Computer Fluidity Groups and Presentation Mode



As described above, a regression analysis was performed in order to control for differences in fluidity that existed among the presentation modes. In addition to fluidity scale score, the regression model included two dummy variables that represented the basic and robust word processing groups. For the first regression model students' topic development score was the dependent variable. As seen in Table 10, the regression model accounted for 10% of the variance in topic development scores. As indicated by the standardized coefficient (Beta), fluidity was the

strongest predictor of students' topic development scores. Both basic and robust word processing conditions had a positive effect on the topic development score, after controlling for differences in fluidity. However, only the coefficient for the robust word processing was statistically significant ($p < .001$).

For the second regression model students' English conventions score was the dependent variable. As seen in Table 10, the regression model accounted for 11% of the variance in English convention scores. As indicated by the standardized coefficient (Beta), fluidity was the strongest predictor of students' writing scores. The basic word processing condition had a negative effect on the English convention score, while the robust word processing condition had a positive effect on the English conventions score, after controlling for differences in fluidity. Both the basic word processing coefficient ($p = .006$) and the robust word processing coefficient ($p < .001$) were found to be statistically significant.

Finally in the third regression model students' total essay score was the dependent variable. As seen in Table 10, the regression model accounted for 11% of the variance in writing scores. As indicated by the standardized coefficient (Beta), fluidity was the strongest predictor of students' writing scores. The basic word processing condition had a negative effect on the English convention score, while the robust word processing condition had a positive effect the English conventions score, after controlling for differences in fluidity. Only the coefficient for the robust word processing was statistically significant ($p < .001$).

(Table 10 is shown on the following page.)

Table 10: Regression Model for Computer Fluidity and Group Membership Predicting Writing Scores

Topic Development					
R ² = .10 F = 29.51 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	5.95	.12		48.08	<.001
Fluidity	.58	.08	.28	7.02	<.001
Basic WP	.04	.21	.01	.19	.85
Robust WP	.63	.17	.14	3.69	<.001
English Standards					
R ² = .11 F = 34.06 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	5.02	.09		59.24	<.001
Fluidity	.38	.06	.27	6.77	<.001
Basic WP	-.39	.14	-.12	2.78	.006
Robust WP	.58	.12	.19	4.97	<.001
Total Score					
R ² = .11 F = 34.28 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	10.97	.19		56.83	<.001
Fluidity	.96	.13	.29	7.47	<.001
Basic WP	-.35	.32	-.05	1.10	.27
Robust WP	1.2	.27	.18	4.55	<.001

Although students were grouped by computer fluidity scores and ANOVAs were performed within each fluidity level to examine differences among composition modes, there remained a statistically significant difference in computer fluidity among composition modes for the high and medium fluidity groups. As seen in Table 11, students in the medium fluidity group who composed their essay on paper had a lower fluidity score than did students in the two computer groups. Students in the high fluidity group who composed their essay using the basic word processing mode had a higher mean fluidity score than did students in the other two groups. For this reason, the regression analyses presented above were performed within the medium and high fluidity group to examine the modal effects while controlling for differences in fluidity scores.

(Table 11 is shown on the following page.)

Table 11: Comparison of Mean Fluidity Scores Among Testing Modes Within Each Fluidity Level

	N	All Students	Paper	Basic WP	Robust WP	F	p
Low	273	-1.11	-1.15	-1.14	-1.02	.80	.45
Medium ^{1,2}	273	.12	.06	.18	.15	6.71	.001
High ^{1,3}	273	.99	.92	1.09	.78	27.89	<.001

1 significant difference between Paper and Basic WP at $p < .05$ using Tukey's post-hoc mean comparison

2 significant difference between Paper and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

3 significant difference between Basic WP and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

As seen in Table 12, the first regression model to the medium fluidity group used students' topic development score as the dependent variable. The regression model accounted for 3% of the variance in topic development scores. As indicated by the standardized coefficient (Beta), the robust word processing condition was the strongest predictor of students' topic development scores. The basic word processing coefficient had a negative effect on topic development scores and the robust word processing condition had a positive effect on the topic development score, after controlling for differences in fluidity. Only the robust word processing ($p = .03$) condition was statistically significant.

In the second regression model for the medium fluidity group, students' English conventions score was the dependent variable. As seen in Table 12, the regression model accounted for 9% of the variance in standard English convention scores. As indicated by the standardized coefficient (Beta), the robust word processing condition was the strongest predictor of students' writing scores. The basic word processing coefficient had a negative effect on standard English conventions scores and the robust word processing condition had a positive effect on the standard English conventions scores, after controlling for differences in fluidity. However only the robust word processing coefficient ($p < .001$) was found to be statistically significant.

Finally, in the third regression model for the medium fluidity group, students' total essay score was the dependent variable. As seen in Table 12, the regression model accounted for 6% of the variance in writing scores. As indicated by the standardized coefficient (Beta), the robust word processing condition was the strongest predictor of students' writing scores. The basic word processing coefficient had a negative effect on writing scores and the robust word processing condition had a positive effect on the writing scores, after controlling for differences in fluidity. However, only the robust word processing coefficient ($p < .001$) was found to be statistically significant.

(Table 12 is shown on the following page.)

Table 12: Regression Model for Computer Fluidity and Group Membership Predicting Writing Scores for Students in the Medium Fluidity Group

Topic Development					
R ² = .03 F = 3.36 p = .02					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	5.86	.21		27.80	<.001
Fluidity	.77	.58	.08	1.33	.18
Basic WP	-.20	.40	-.03	.50	.62
Robust WP	.65	.29	.15	2.25	.03
English Standards					
R ² = .09 F = 10.28 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	4.82	.13		36.35	<.001
Fluidity	.63	.36	.10	1.73	.09
Basic WP	-.44	.25	-.11	-1.72	.09
Robust WP	.67	.18	.24	3.73	<.001
Total Score					
R ² = .06 F = 6.40 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	10.68	.32		33.48	<.001
Fluidity	1.40	.88	.10	1.60	.11
Basic WP	-.64	.61	-.07	1.04	.30
Robust WP	1.32	.43	.20	3.04	.003

As seen in Table 13, the first regression model for the high fluidity group uses students' topic development score as the dependent variable. The regression model accounted for 6% of the variance in topic development scores. As indicated by the standardized coefficient (Beta), the robust word processing condition was the strongest predictor of students' topic development scores. Both the basic and the robust word processing condition have a positive effect on the topic development score, after controlling for differences in fluidity. Only the robust word processing (p=.004) condition was statistically significant.

In the second regression model for the high fluidity group, students' English conventions score was the dependent variable. As seen in Table 13, the regression model accounted for 18% of the variance in standard English convention scores. As indicated by the standardized coefficient (Beta), the robust word processing condition was the strongest predictor of students' writing scores. The basic word processing coefficient had a negative effect on standard English conventions scores and the robust word processing condition had a positive effect on the English

conventions scores, after controlling for differences in fluidity. However only the robust word processing coefficient ($p < .001$) was statistically significant.

Finally, in the third regression model for the high fluidity group, students' total essay score was the dependent variable. As seen in Table 13, the regression model accounted for 12% of the variance in writing scores. As indicated by the standardized coefficient (Beta), the robust word processing condition was the strongest predictor of students' writing scores. Both the basic and the robust word processing condition have a positive effect on the writing score, after controlling for differences in fluidity. However only the robust word processing coefficient ($p < .001$) was statistically significant.

Table 13: Regression Model for Computer Fluidity and Group Membership Predicting Writing Scores for Students in the High Fluidity Group

Topic Development					
R ² = .06 F = 6.39 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	5.29	.51		10.33	<.001
Fluidity	1.06	.37	.19	2.89	.004
Basic WP	.23	.41	.06	.55	.58
Robust WP	1.27	.44	.31	2.89	.004
English Standards					
R ² = .18 F = 21.07 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	4.04	.37		10.93	<.001
Fluidity	1.10	.26	.25	4.15	<.001
Basic WP	-.18	.30	-.06	.61	.55
Robust WP	1.34	.32	.42	4.24	<.001
Total Score					
R ² = .12 F = 13.35 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	9.33	.81		11.53	<.001
Fluidity	2.16	.58	.23	3.73	<.001
Basic WP	.05	.65	.01	.07	.94
Robust WP	2.61	.70	.39	3.76	<.001

Computer Literacy Analysis

As previously described, students were assigned to a high, medium, or low computer literacy group based on their computer literacy scaled score. Table 14 displays the average topic development, English conventions, and total writing scores across the three presentation modes for students based on their computer literacy categorization.

Table 14: Comparison of Mean Writing Test Score Across Computer Literacy Groups and Presentation Mode

	N	All Students	Paper	Basic WP	Robust WP	F	p
Topic Development							
Low	261	5.06	4.87	5.15	5.30	1.27	.28
Medium ²	264	6.07	5.75	6.05	6.47	3.77	.02
High ^{1,2}	260	7.13	6.38	7.29	7.63	10.16	<.001
English Standards							
Low ³	261	4.41	4.45	4.04	4.74	4.98	.008
Medium ^{2,3}	264	4.98	4.63	4.72	5.61	17.66	<.001
High ^{2,3}	260	5.78	5.37	5.63	6.23	11.18	<.001
Total Score							
Low	261	9.48	9.32	9.19	10.04	1.81	.17
Medium ^{2,3}	264	11.05	10.37	10.77	12.08	9.52	<.001
High ^{2,3}	260	12.90	11.75	12.92	13.86	11.61	<.001

1 significant difference between Paper and Basic WP at $p < .05$ using Tukey's post-hoc mean comparison

2 significant difference between Paper and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

3 significant difference between Basic WP and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

In general, students with higher computer literacy scores tended to score higher on the writing assessment across all three modes. Similarly, students in the low computer literacy group performed lower, on average, across each of the presentation modes. Since this relationship holds across all three testing modes, it does not provide evidence that students taking the assessment using the two computer-based modes were advantaged or disadvantaged by their computer literacy.

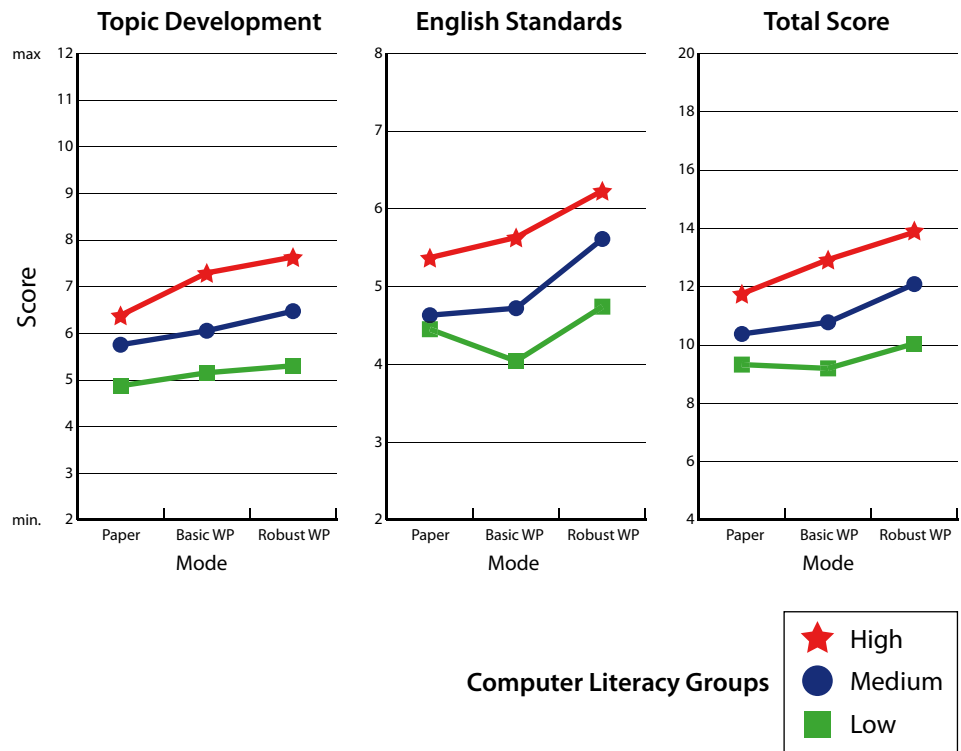
To examine differences within each computer literacy group across presentation modes, a one-way ANOVA was run across presentation modes for each literacy categorization. For the topic development scores, there were no significant differences in scores at the low level, but there were statistically significant differences in scores at the medium ($p = .02$) and high ($p < .001$) levels. At the medium level, the differences were found between the paper group and the robust word processing group. At the high level, the differences were found between paper and basic word processing and between the paper and robust word processing

groups. In both cases the robust word processing group achieved the higher average score than the paper group and at the high level, the basic word processing group achieved a higher average score than the paper group.

For the standard English conventions scores, there were significant differences in scores at the low ($p=.008$), medium ($p<.001$), and high ($p<.001$) levels. At the low level the difference was found between the basic word processing group and the robust word processing group. At both the medium and high levels, the differences were found between paper and robust word processing and between the basic word processing and robust word processing. In each case, the robust word processing group achieved the higher average score.

For the total writing scores, there were no significant differences in scores at the low level, but there were statistically significant differences in scores at the medium ($p<.001$) and high ($p<.001$) levels. At both the medium and high levels, the differences were found between paper and robust word processing and between the basic word processing and robust word processing. In each case, the robust word processing group achieved the higher average score.

Figure 5: Mean Writing Scores for Computer Literacy Groups and Presentation Mode



To control for differences in computer literacy that existed among the presentation modes, a regression analysis was performed. In addition to computer literacy score, the regression model included two dummy variables that represented the basic and robust word processing groups. For the first regression model

students' topic development score was the dependent variable. As seen in Table 15, the regression model accounted for 20% of the variance in topic development scores. As indicated by the standardized coefficient (Beta), computer literacy was the strongest predictor of students' topic development scores. Both basic and robust word processing conditions had a positive effect on the topic development score, after controlling for differences in computer literacy. The coefficient for the basic word processing group was statistically significant at the $p=.003$ level, while the coefficient for the robust word processing group was statistically significant at the $p<.001$ level.

For the second regression model students' English conventions score was the dependent variable. As seen in Table 15, the regression model accounted for 21% of the variance in standard English convention scores. As indicated by the standardized coefficient (Beta), computer literacy was the strongest predictor of students' writing scores. The basic word processing condition had a negative effect on the standard English convention score, while the robust word processing condition had a positive effect on the standard English conventions score, after controlling for differences in computer literacy. Only the robust word processing coefficient ($p<.001$) was statistically significant.

Finally, in the third regression model, students' total essay score was the dependent variable. As seen in Table 15, the regression model accounted for 23% of the variance in writing scores. As indicated by the standardized coefficient (Beta), computer literacy was the strongest predictor of students' writing scores. Both basic and robust word processing conditions had a positive effect on the topic development score, after controlling for differences in computer literacy. However, only the coefficient for the robust word processing was statistically significant ($p<.001$).

(Table 15 is shown on the following page.)

Table 15: Regression Model for Computer Literacy and Group Membership Predicting Writing Scores

Topic Development					
R ² = .20 F = 65.48 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	5.68	.11		53.13	<.001
Literacy	.83	.07	.40	12.36	<.001
Basic WP	.48	.17	.11	2.93	.003
Robust WP	.81	.16	.18	5.11	<.001
English Standards					
R ² = .21 F = 70.19 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	4.83	.07		66.25	<.001
Literacy	.53	.05	.37	11.66	<.001
Basic WP	-.03	.11	-.01	.28	.78
Robust WP	.71	.11	.23	6.58	<.001
Total Score					
R ² = .23 F = 78.57 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	10.51	.17		63.92	<.001
Literacy	1.36	.10	.42	13.21	<.001
Basic WP	.45	.25	.06	1.78	.08
Robust WP	1.53	.24	.22	6.24	<.001

To examine differences among composition modes, ANOVAs were performed within each computer literacy level. As seen in Table 16, students did not differ significantly in their computer literacy skills across delivery modes at each literacy level. Since there were no statistically significant differences in computer literacy between testing modes within each level of computer literacy, no further analyses were conducted within the computer literacy groups.

Table 16: Comparison of Mean Computer Literacy Among Testing Modes Within Each Literacy Level

	N	All Students	Paper	Basic WP	Robust WP	F	p
Low	261	-1.16	-1.20	-1.15	-1.09	.37	.69
Medium	264	.25	.27	.23	.24	.85	.43
High	260	.91	.91	.92	.89	.43	.65

Computer Use for Writing Analysis

Students' scale scores for using computers for writing (writing use) were used to divide students into three groups which represented high, medium, and low levels of reported use of computers for writing. Table 17 displays the average writing scores across the three presentation modes for students based on their writing use categorization.

Table 17: Comparison of Mean Writing Test Score Across Writing Use Groups and Presentation Mode

	N	All Students	Paper	Basic WP	Robust WP	F	p
Topic Development							
Low	194	6.00	5.69	5.88	6.52	2.64	.07
Medium ^{1,2}	191	6.14	5.39	6.75	6.42	7.60	.001
High ^{1,2}	304	5.92	5.32	6.19	6.36	9.47	<.001
English Standards							
Low ^{2,3}	194	5.09	4.87	4.79	5.65	7.27	.001
Medium ¹	191	5.12	4.73	5.09	5.57	5.94	.003
High ^{2,3}	304	4.88	4.48	4.72	5.42	14.00	<.001
Total Score							
Low ^{2,3}	194	11.09	10.56	10.67	12.17	4.60	.01
Medium ^{1,2}	191	11.26	10.11	11.84	11.98	7.05	.001
High ^{1,2}	304	10.80	9.80	10.92	11.78	12.17	<.001

1 significant difference between Paper and Basic WP at $p < .05$ using Tukey's post-hoc mean comparison

2 significant difference between Paper and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

3 significant difference between Basic WP and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

There is not a clear pattern between students reported use of computers for writing and their score on the writing assessment. Overall, students with high reported use of computers for writing performed lower than students who reportedly use computers for writing less often.

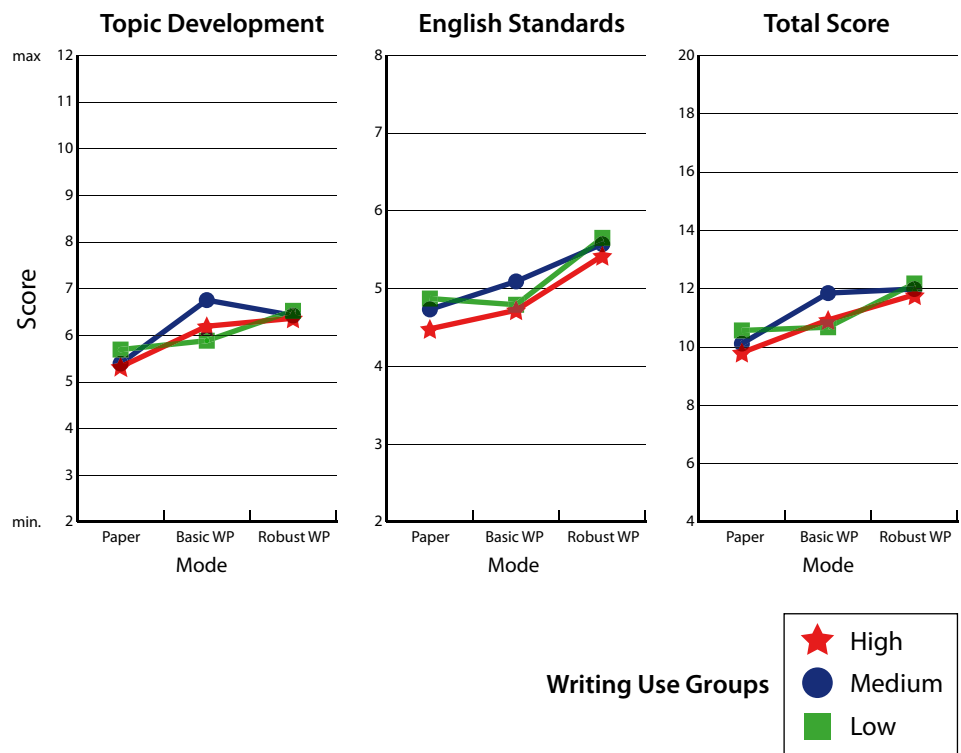
To examine differences within each writing use group across presentation modes, a one-way ANOVA was run across presentation modes for each writing use categorization. For the topic development scores, there were no significant differences in scores at the low level, but there were statistically significant differences in scores at the medium ($p = .001$) and high ($p < .001$) levels. At both the medium and high levels, the differences were found between the paper group and the basic word processing group and between the paper and robust word processing groups. In both cases the word processing groups achieved the higher average score than the paper group.

For the standard English conventions scores, there were significant differences in scores at the low ($p = .001$), medium ($p = .003$), and high ($p < .001$) levels. At the

low and high levels the difference were found between the paper and robust word processing groups and between the basic and robust word processing groups. In both cases the robust word processing group achieved a higher average score than the comparison group. At the medium level, the differences were found between paper and basic word processing groups and the paper and robust word processing groups. At this level, the basic and robust word processing groups achieved higher average scores than the paper group.

For the total writing scores, there were significant differences in scores at the low ($p=.01$), medium ($p=.001$), and high ($p<.001$) levels. At the low level there were differences between the paper and robust word processing group and between the basic and robust word processing groups. In both cases the robust word processing group achieved a higher average score than the comparison group. At both the medium and high levels, the differences were found between paper and basic word processing and between the paper and robust word processing groups. In each case, the word processing groups achieved the higher average score.

Figure 6: Mean Writing Scores for Writing Use Groups and Presentation Mode



To control for differences in writing use that existed among the presentation modes, a regression analysis was performed. In addition to writing use score, the regression model included two dummy variables that represented the basic and robust word processing groups. For the first regression model students' topic development score was the dependent variable. As seen in Table 18, the regression model accounted for 5% of the variance in topic development scores. As

indicated by the standardized coefficient (Beta), robust word processing group was the strongest predictor of students' writing scores. Both basic and robust word processing conditions had a positive effect on the topic development score, after controlling for differences in writing use. The coefficient for the basic and robust word processing groups were statistically significant at the $p < .001$ level.

For the second regression model students' English conventions score was the dependent variable. As seen in Table 18, the regression model accounted for 8% of the variance in standard English convention scores. As indicated by the standardized coefficient (Beta), the robust word processing group was the strongest predictor of students' writing scores. Both the basic and robust word processing conditions had a positive effect on the standard English convention score, after controlling for differences in writing use. Only the robust word processing coefficient ($p < .001$) was found to be statistically significant.

Finally, in the third regression model students' total essay score was the dependent variable. As seen in Table 18, the regression model accounted for 6% of the variance in writing scores. As indicated by the standardized coefficient (Beta), the robust word processing group was the strongest predictor of students' writing scores. Both the basic and robust word processing conditions had a positive effect on the total writing score, after controlling for differences in writing use. The coefficient for the basic word processing group was significant at the $p = .001$ level while the coefficient for robust word processing was significant at the $p < .001$ level.

(Table 18 is shown on the following page.)

Table 18: Regression Model for Writing Use and Group Membership Predicting Writing Scores

Topic Development					
R ² = .05 F = 12.28 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	5.45	.12		43.77	<.001
Writing Use	-.15	.08	-.07	1.99	.05
Basic WP	.81	.19	.18	4.24	<.001
Robust WP	.97	.18	.22	5.33	<.001
English Standards					
R ² = .08 F = 19.95 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	4.66	.09		54.77	<.001
Writing Use	-.15	.05	-.10	2.85	.005
Basic WP	.18	.13	.06	1.38	.17
Robust WP	.86	.12	.29	6.95	<.001
Total Score					
R ² = .06 F = 16.17 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	10.11	.19		52.12	<.001
Writing Use	-.30	.12	-.09	-2.53	.01
Basic WP	.99	.30	.14	3.33	.001
Robust WP	1.83	.28	.27	6.47	<.001

Although students were grouped by writing use score and ANOVAs were performed within each writing use level to examine differences among composition modes, there remained a statistically significant difference in writing use among composition modes for the medium writing use group. As seen in Table 19, students in the medium writing use group who composed their essay on paper had higher writing use than did students in the two computer groups. For this reason, the regression analyses presented above were performed within the medium writing use group to examine the modal effects while controlling for differences in writing use.

Table 19: Comparison of Mean Writing Use Among Testing Modes Within Each Writing Use Level

	N	All Students	Paper	Basic WP	Robust WP	F	P
Low	194	-1.02	-1.06	-1.00	-1.00	1.48	.23
Medium	191	-.34	-.32	-.35	-.36	3.18	.04
High	304	.87	.93	.84	.83	.47	.63

As seen in Table 20, the first medium writing use regression model accounted for 7% of the variance in topic development scores. As indicated by the standardized coefficient (Beta), the basic word processing group was the strongest predictor of students' writing scores. Both basic and robust word processing conditions had a positive effect on the topic development score, after controlling for differences in writing use. The coefficient for the basic and robust word processing groups were statistically significant at the $p=.001$ and $p=.01$ levels respectively.

For the second medium writing use regression model, students' English conventions score was the dependent variable. As seen in Table 20, the regression model accounted for 5% of the variance in standard English convention scores. As indicated by the standardized coefficient (Beta), the robust word processing group was the strongest predictor of students' writing scores. Both the basic and robust word processing conditions had a positive effect on the standard English convention score, after controlling for differences in writing use. Only the robust word processing coefficient ($p=.002$) was found to be statistically significant.

Finally in the third medium writing uses regression model students' total essay score was the dependent variable. As seen in Table 20, the regression model accounted for 7% of the variance in writing scores. As indicated by the standardized coefficient (Beta), the robust word processing group was the strongest predictor of students' writing scores. Both the basic and robust word processing conditions had a positive effect on the total writing score, after controlling for differences in writing use. The coefficient for the basic word processing group was significant at the $p=.006$ level while the coefficient for the robust word processing was significant at the $p=.003$ level.

Table 20: Regression Model for Writing Use and Group Membership Predicting Writing Scores for Students in the Medium Writing Use Group

Topic Development					
R ² = .07 F = 6.03 p = .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	4.66	.50		9.23	<.001
Writing Use	-2.28	1.38	-.12	1.66	.10
Basic WP	1.29	.37	.27	3.45	.001
Robust WP	.93	.36	.21	2.56	.01
English Standards					
R ² = .05 F = 4.58 p = .004					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	4.32	.35		12.52	<.001
Writing Use	-1.27	.94	-.10	1.35	.18
Basic WP	.32	.26	.10	1.25	.22
Robust WP	.78	.25	.26	3.16	.002
Total Score					
R ² = .07 F = 5.67 p = .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	8.98	.78		11.54	<.001
Writing Use	-3.56	2.13	-.12	1.67	.10
Basic WP	1.61	.58	.22	2.79	.006
Robust WP	1.71	.56	.25	3.06	.003

School Use Analysis

Students' scaled school use scores were used to divide students into three groups that represented high, medium, and low levels of reported use of computers in school for academic purposes. Table 21 displays the average writing scores across the three composition modes for students based on their school computer use categorization.

Table 21: Comparison of Mean Writing Test Score Across School Computer Use Groups and Presentation Mode

	N	All Students	Paper	Basic WP	Robust WP	F	p
Topic Development							
Low ²	226	6.03	5.54	6.25	6.43	3.75	.03
Medium ²	227	6.31	5.70	6.45	6.76	5.99	.003
High ¹	224	5.69	5.21	6.13	5.91	5.57	.004
English Standards							
Low ^{2,3}	226	5.13	4.91	4.93	5.61	5.93	.003
Medium ^{2,3}	227	5.10	4.64	4.87	5.63	12.22	<.001
High ²	224	4.82	4.50	4.81	5.28	6.20	.002
Total Score							
Low ²	226	11.16	10.45	11.17	12.04	4.49	.01
Medium ²	227	11.41	10.34	11.33	12.39	9.13	<.001
High ^{1,2}	224	10.50	9.71	10.94	11.19	5.72	.004

1 significant difference between Paper and Basic WP at $p < .05$ using Tukey's post-hoc mean comparison

2 significant difference between Paper and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

3 significant difference between Basic WP and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

There is not a clear pattern between students' reported school computer use and their score on the writing assessment. Overall, students with high reported school computer use performed slightly lower than students who reportedly use computers in school less often.

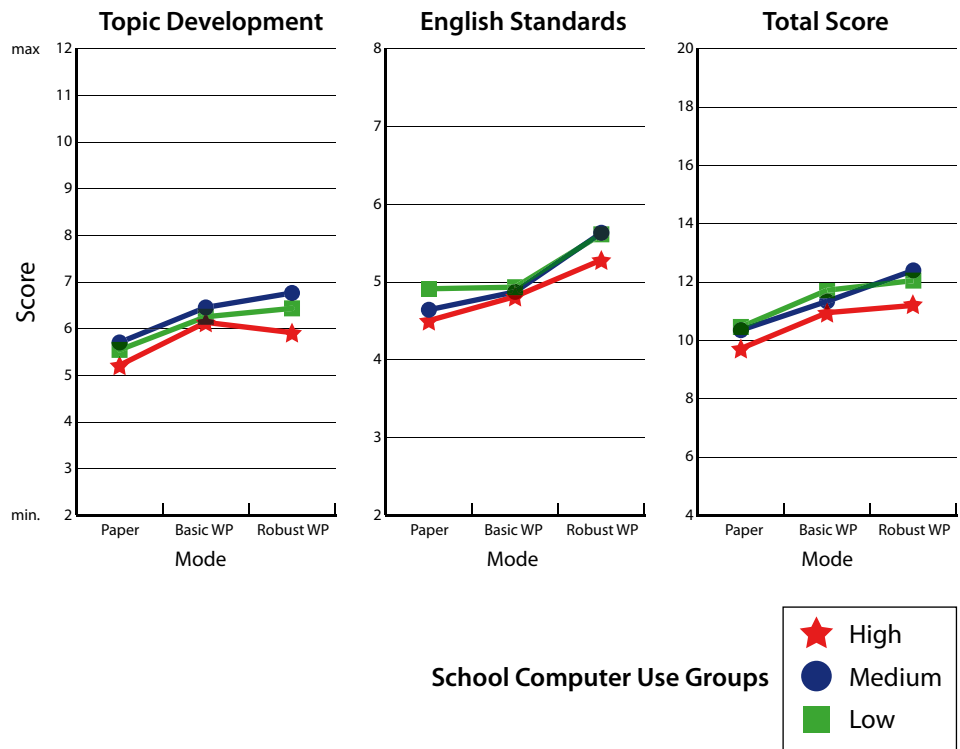
To examine differences within each school computer use group across presentation modes, a one-way ANOVA was run across presentation modes for each school computer use categorization. For the topic development scores, there were significant differences in scores at the low ($p = .03$), medium ($p = .003$), and high ($p = .004$) levels. At both the low and medium levels, the differences were found between the paper group and the robust word processing group. At the high level the differences were found between the paper and basic word processing groups. In all cases the word processing group achieved a higher average score than the paper group.

For the standard English conventions scores, there were significant differences in scores at the low ($p = .003$), medium ($p < .001$), and high ($p = .002$) levels. At the

low and medium levels differences were found between the paper and robust word processing groups and between the basic and robust word processing groups. In both cases the robust word processing group achieved a higher average score than the comparison group. At the high level, the differences were found between paper and robust word processing groups. In this case the robust word processing group achieved a higher average score than the paper group.

For the total writing scores, there were significant differences in scores at the low ($p=.01$), medium ($p<.001$), and high ($p=.004$) levels. At the low and medium levels there were differences between the paper and robust word processing group. In both cases the robust word processing group achieved a higher average score than the paper group. At the high level, the differences were found between paper and basic word processing and between the paper and robust word processing groups. In each case, the word processing group achieved a higher average score than the paper group.

Figure 7: Mean Writing Scores for School Use Groups and Presentation Mode



To control for differences in school computer use that existed among the presentation modes, a regression analysis was performed. In addition to school computer use score, the regression model included two dummy variables that represented the basic and robust word processing groups. For the first regression model students' topic development score was the dependent variable. As seen in Table 22, the regression model accounted for 4% of the variance in topic development scores. As indicated by the standardized coefficient (Beta), the robust word processing group was the strongest predictor of students' topic development scores. Both basic and robust word processing conditions had a positive effect on the topic development score, after controlling for differences in school computer use. The coefficient for the basic and robust word processing groups were statistically significant at the $p < .001$ level.

For the second regression model, students' English conventions score was the dependent variable. As seen in Table 22, the regression model accounted for 7% of the variance in standard English convention scores. As indicated by the standardized coefficient (Beta), the robust word processing group was the strongest predictor of students' standard English conventions scores. Both the basic and robust word processing conditions had a positive effect on the standard English convention score, after controlling for differences in school computer use. Only the robust word processing coefficient ($p < .001$) was found to be statistically significant.

Finally in the third regression model students' total essay score was the dependent variable. As seen in Table 22, the regression model accounted for 6% of the variance in writing scores. As indicated by the standardized coefficient (Beta), the robust word processing group was the strongest predictor of students' writing scores. Both the basic and robust word processing conditions had a positive effect on the total writing score, while the school computer use score had a negative effect the writing score. The coefficient for the basic word processing group was significant at the $p = .001$ level while the coefficient for the robust word processing group was significant at the $p < .001$ level.

(Table 22 is shown on the following page.)

Table 22: Regression Model for Keyboarding Speed and Group Membership Predicting Reading Scores

Topic Development					
R ² = .04 F = 11.44 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	5.47	.13		43.56	<.001
School Use	-.15	.08	-.07	1.93	.05
Basic WP	.79	.19	.17	4.11	<.001
Robust WP	.94	.18	.22	5.11	<.001
English Standards					
R ² = .7 F = 17.89 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	4.68	.09		54.69	<.001
School Use	-.12	.05	-.08	2.24	.03
Basic WP	.19	.13	.06	1.43	.15
Robust WP	.84	.13	.28	6.73	<.001
Total Score					
R ² = .06 F = 14.79 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	10.16	.20		51.91	<.001
School Use	-.27	.12	-.08	2.23	.03
Basic WP	.98	.30	.14	3.27	.001
Robust WP	1.78	.29	.26	6.23	<.001

ANOVAs were performed within each school computer use level to examine differences among composition modes. As seen in Table 23, students did not differ significantly in their reported school computer use across delivery modes at each school computer use level. For this reason, no further analyses were performed based on students' school use of computers.

Table 23: Comparison of Mean School Computer Use among Testing Modes within Each School Computer Use Level

	N	All Students	Paper	Basic WP	Robust WP	F	P
Low	226	-.90	-.92	-.91	-.88	.43	.65
Medium	227	-.21	-.21	-.19	-.22	.48	.62
High	224	1.12	1.10	1.06	1.23	.69	.50

Home Use Analysis

Students' scaled school use scores were used to divide students into three groups that represented high, medium, and low levels of reported use of computers at home. Table 24 displays the average writing scores across the three composition modes for students based on their home computer use categorization.

Table 24: Comparison of Mean Writing Test Score across Home Computer Use Groups and Presentation Mode

	N	All Students	Paper	Basic WP	Robust WP	F	P
Topic Development							
Low	227	5.84	5.59	6.05	5.93	.99	.37
Medium ^{1,2}	226	6.17	5.42	6.60	6.62	9.91	<.001
High ^{1,2}	228	6.05	5.38	6.28	6.75	9.61	<.001
English Standards							
Low ²	227	4.95	4.72	4.82	5.32	3.86	.02
Medium ^{2,3}	226	5.21	4.77	5.11	5.71	10.16	<.001
High ^{2,3}	228	4.85	4.53	4.71	5.45	8.62	<.001
Total Score							
Low	227	10.78	10.31	10.86	11.25	1.67	.19
Medium ^{1,2}	226	11.39	10.19	11.70	12.33	10.91	<.001
High ²	228	10.89	9.91	10.99	12.20	9.81	<.001

1 significant difference between Paper and Basic WP at $p < .05$ using Tukey's post-hoc mean comparison

2 significant difference between Paper and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

3 significant difference between Basic WP and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

Generally, students with medium levels of home computer use had higher scores for all three writing scores than did students in the low or high levels of home computer use.

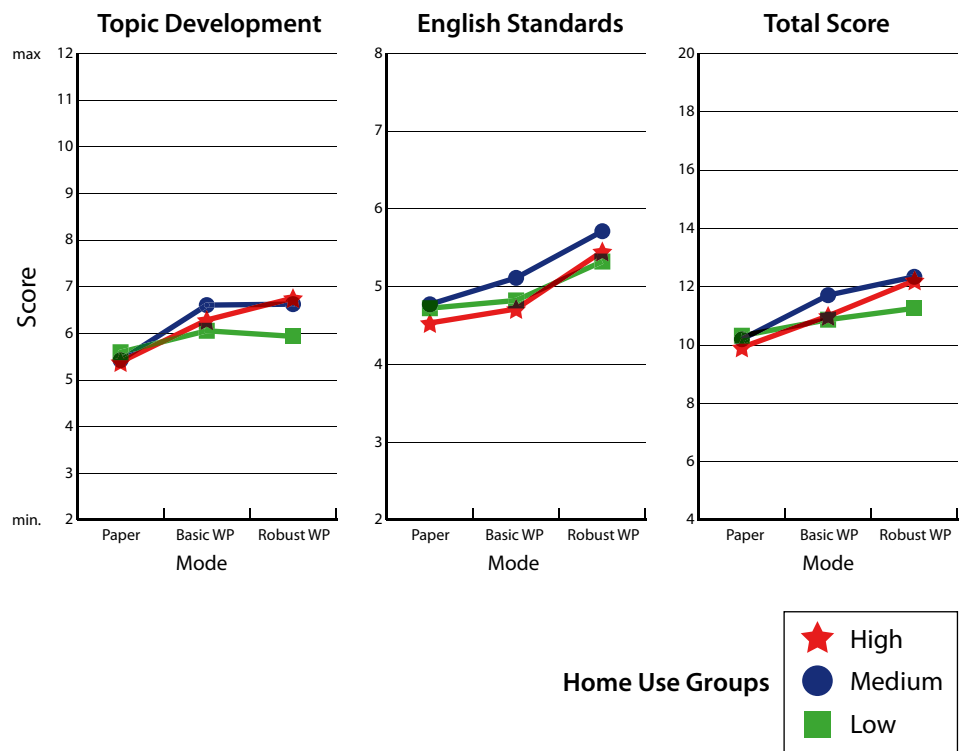
To examine differences within each home computer use group across presentation modes, a one-way ANOVA was run across presentation modes for each home computer use categorization. For the topic development scores, there were no significant differences at the low level, but there were significant differences in scores at the medium ($p = .003$) and high ($p = .004$) levels. At both the medium and high levels, the differences were found between the paper group and the robust word processing group and between the basic and robust word processing groups. In both cases the robust word processing group achieved a higher average score than the paper group.

For the English conventions scores, there were significant differences in scores at the low ($p = .02$), medium ($p < .001$), and high ($p < .001$) levels. At the low level there were differences between the paper and robust word processing groups. At the medium and high levels the differences were found between the paper and

robust word processing groups and between the basic and robust word processing groups. In all cases the robust word processing group achieved a higher average score than the other groups.

For the total writing scores, there were no significant differences in scores at the low level ($p=.19$), but there were significant differences in scores at the medium ($p<.001$) and high ($p<.001$) levels. At the medium level there were differences between the paper and basic word processing group and between the paper and robust word processing groups. At the high level, the differences were found between paper and basic word processing groups. In all cases the word processing groups achieved a higher average score than the paper group.

Figure 8: Mean Writing Scores for Home Use Groups and Presentation Mode



To control for differences in home computer use that existed among the presentation modes, a regression analysis was performed. In addition to home computer use score, the regression model included two dummy variables which represented the basic and robust word processing groups. For the first regression model students' topic development score was the dependent variable. As seen in Table 25, the regression model accounted for 5% of the variance in topic development scores. As indicated by the standardized coefficient (Beta), the robust word processing group was the strongest predictor of students' topic development scores. Both basic and robust word processing conditions had a positive effect on the topic development score, after controlling for differences in home computer use.

The coefficients for the basic and robust word processing groups were statistically significant at the $p < .001$ level.

For the second regression model, students' English conventions score was the dependent variable. As seen in Table 25, the regression model accounted for 6% of the variance in standard English convention scores. As indicated by the standardized coefficient (Beta), the robust word processing group was the strongest predictor of students' standard English conventions scores. Both the basic and robust word processing conditions had a positive effect on the standard English convention score, after controlling for differences in home computer use. Only the robust word processing coefficient ($p < .001$) was found to be statistically significant.

Finally in the third regression model students' total essay score was the dependent variable. As seen in Table 25, the regression model accounted for 5% of the variance in writing scores. As indicated by the standardized coefficient (Beta), the robust word processing group was the strongest predictor of students' writing scores. Both the basic and robust word processing conditions had a positive effect on the total writing score, after controlling for differences in home computer use. The coefficient for the basic word processing group was significant at the $p = .001$ level while the coefficient for the robust word processing was significant at the $p < .001$ level.

(Table 25 is shown on the following page.)

Table 25: Regression Model for Home Use and Group Membership Predicting Reading Scores

Topic Development					
R ² = .05 F = 12.55 p < .001					
Variable	Coefficient	SE	Beta	T ratio	Sig.
Intercept	5.46	.13		43.71	<.001
Home Use	.17	.08	.08	2.22	.03
Basic WP	.83	.19	.18	4.35	<.001
Robust WP	.98	.18	.22	5.33	<.001
English Standards					
R ² = .06 F = 15.23 p < .001					
Variable	Coefficient	SE	Beta	T ratio	Sig.
Intercept	4.67	.09		53.57	<.001
Home Use	.03	.05	.02	.50	.62
Basic WP	.19	.13	.06	1.44	.15
Robust WP	.84	.13	.27	6.55	<.001
Total Score					
R ² = .05 F = 14.00 p < .001					
Variable	Coefficient	SE	Beta	T ratio	Sig.
Intercept	10.13	.20		51.47	<.001
Home Use	.20	.12	.06	1.63	.10
Basic WP	1.02	.30	.14	3.40	.001
Robust WP	1.81	.29	.26	6.28	<.001

To examine differences among composition modes, ANOVAs were performed within each home computer use level. As seen in Table 26, students did not differ significantly in their reported home computer use across delivery modes at each home computer use level. For this reason, no further analyses were conducted based on students' home computer use.

Table 26: Comparison of Mean Keyboard Speed Among Testing Modes within Each Keyboard Level

	N	All Students	Paper	Basic WP	Robust WP	F	p
Low	227	6.17	5.42	6.60	6.62	.03	.97
Medium	226	.08	.04	.12	.10	1.57	.21
High	228	1.06	1.08	1.04	1.06	.17	.84

Examining the Modal Effect by Demographic Variables

As noted earlier, the modal effect documented in previous studies and subsequent recommendations to allow some students to compose essay responses on a computer has sparked concerns about equity. In most cases, these concerns focus on students in urban and suburban schools and result from a belief that students in wealthier suburban schools will benefit more by being allowed to use a computer for writing tests. To examine whether the magnitude of the modal effect varies by student background, a series of analyses was conducted using demographic information provided by students through the student survey. Specifically, the background variables examined include: Gender, Race/Ethnicity, multilingual status, and IEP status. For each background variable, total group differences were first examined for each of the writing scores and for the computer fluidity measure. In addition, within each category of the background variable, the modal effect was examined using an ANOVA. When differences among modal groups (i.e., paper, basic word processor, and robust word processor) were detected, post hoc analysis were conducted to identify between which groups the effect was statistically significant. To compare the magnitude of the effect among the categories for each background variable and to examine the extent to which the modal effect interacted with each background variable, an interaction graph was created. Finally, to test for statistically significant interactions, a series of regression analyses was performed in which interaction terms were included. The method used to form these interaction terms are explained below.

To classify students into categories for each background variable, student responses to four items on the student survey were used.

- Students were classified as male or female based on their response to the following item, “Are you a boy or a girl?” Students responded either “boy” or “girl.”
- Students were classified as white or non-white based on their response to the following item, “Which best describes you? Please mark as many as apply.” Students were given six response options. As a result, more than 20 combinations were marked, many of which contained a small number of students. Given that concerns regarding the modal effect generally focus on differential effects for white versus non-white students, students were collapsed into one of two race/ethnicity categories, white or non-white. To do so, all students who responded only that they were White were grouped into the white category while all students who marked any option other than or in addition to White were grouped into the non-white category.
- Students were classified as Multilingual or Non-Multilingual based on their response to the following item, “What language other than English do you speak at home?” Students responded that they did not speak any other language or listed one or more languages. To be clear,

many students who were classified as Multilingual have never been or may no longer be classified as ELL students. Thus, it is important not to confuse the multilingual analysis classification with an ELL classification. Since we were unable to obtain ELL status information for all students from their schools, we employ the multilingual status as a crude approximation (at best) to examine the interaction between the modal effect and language status.

- Students were classified as IEP or non-IEP based on their response to the following item, “Do you have an Individual Education Plan (IEP)?” Students responded “yes,” “no,” or “I don’t know.” During pilot testing, many students who did not have an IEP indicated that they did not know what an IEP was. To avoid multiple students from asking what an IEP is and, in turn, potentially identifying a student who has an IEP, we included the “I don’t know” option. For the analyses presented below, it was assumed that students that responded “I don’t know” did not have an IEP. Thus, to form the non-IEP group, students whom responded “No” or “I don’t know” were combined into a single group.

Modal Effect by Gender

Of the 706 students for whom gender information is available, 51.2% indicated that they were male. As seen in Table 27, females received significantly higher Topic Development, standard English conventions, and total writing scores. In addition, females were significantly faster keyboarders. However, there were not statistically significant differences in computer fluidity, computer literacy or use of computers for writing between males and females.

Table 27: Comparison of Writing and Computer Use Measures by Gender

	Male (357)		Female (349)		p
	Mean	St. Dev.	Mean	St. Dev.	
Topic Development	5.62	1.98	6.39	2.09	<.001
English Standards	4.78	1.39	5.20	1.47	<.001
Total Score	10.40	3.09	11.59	3.32	<.001
Keyboard Speed	111.50	56.14	131.91	54.27	<.001
Fluidity	.11	.91	-.01	1.02	.12
Literacy	.00	1.04	.10	.91	.21
Writing Use	.00	.96	.00	1.04	.99

To examine the extent to which the mode of administration affected the performance of males and females, separate ANOVAs were performed within each gender category. As seen in Table 28, for both males and females, there were significant differences among scores across administration modes. For all writing scales, students in the paper group performed lower than students in either computer

group. Similarly, for all writing scales, students in the robust word processor group received higher scores than students in the basic word processor group. For the topic development scale, for both males and females, students in the paper group performed significantly lower than students in either word processing group, but score differences were not significant between the two word processing groups. For the standard English conventions scale, students in the robust word processor group performed significantly higher than students in either the paper or basic word processor group. Scores between the paper and word processor group did not differ significantly on the standard English conventions scale. For the total score, for both males and females, students in the robust word processor group performed significantly better than the paper group. For females, students in the robust word processor group also performed significantly better than the basic word processor group, while the differences between the basic word processor and the paper group were not significantly different. Conversely, for boys, there was a statistically significant difference between the paper and basic word processor group, but not between the basic and robust word processor groups.

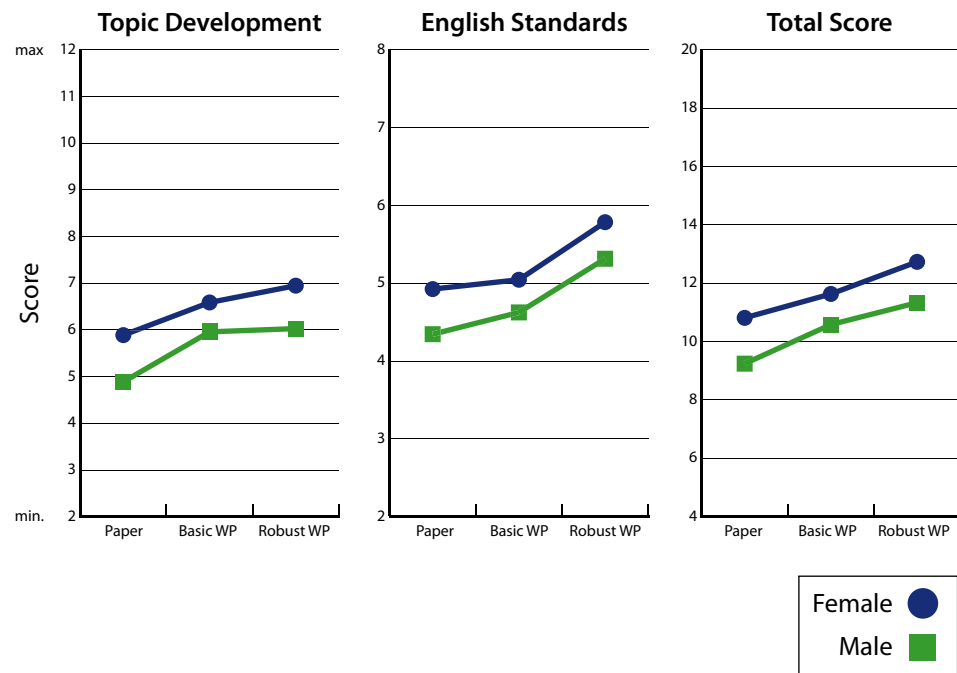
Table 28: Comparison of Mean Writing Test Score Across Presentation Mode by Gender

	N	All Students	Paper	Basic WP	Robust WP	F	p
Topic Development							
Male ^{1,2}	357	5.62	4.88	5.95	6.02	13.27	<.001
Female ^{1,2}	349	6.39	5.88	6.58	6.94	8.53	<.001
English Standards							
Male ^{2,3}	357	4.78	4.34	4.62	5.31	17.69	<.001
Female ^{2,3}	349	5.20	4.92	5.04	5.78	11.85	<.001
Total Score							
Male ^{1,2}	357	10.40	9.23	10.57	11.32	15.90	<.001
Female ^{2,3}	349	11.59	10.80	11.62	12.72	10.47	<.001

1 significant difference between Paper and Basic WP at $p < .05$ using Tukey's post-hoc mean comparison

2 significant difference between Paper and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

3 significant difference between Basic WP and Robust WP at $p < .05$ using Tukey's post-hoc mean comparison

Figure 9: Mean Writing Scores for Gender by Administration Mode

Although the graphs shown in Figure 9 suggest that there is not an interaction between gender and the mode of administration effect, a series of regression analyses were conducted in which each writing scale served as the dependent variable while computer fluidity, gender, and word processing group membership served as the predictors. In addition, an interaction term for each word processing group was also included in each model. The interaction term was created by multiplying gender status (coded 1 for female) and word processor group status. Thus the interaction terms represented the effect of females composing responses on a basic word processor and on a robust word processor. As seen in Table 29, computer fluidity, gender, and the robust word processing condition were significant predictors of each of the writing scale scores. The basic word processing condition was a significant predictor for topic development, but not the two other writing scales. Finally, the interactions between gender and the two word processing conditions were not significant predictors for any of the writing scales.

(Table 29 is shown on the following page.)

Table 29: Regression Model for Computer Fluidity, Gender, Group Membership and Gender-by-Administration Mode Predicting Writing Scores

Topic Development					
R ² = .16 F = 20.22 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	5.15	.19		27.91	<.001
Fluidity	.64	.09	.30	6.95	<.001
Gender	1.20	.25	.29	4.83	<.001
Basic WP	.58	.29	.13	2.00	.05
Robust WP	.88	.25	.21	3.55	<.001
Gender x BWP	-.71	.38	-.12	1.88	.061
Gender x RWP	-.23	.36	-.04	0.64	.53
English Standards					
R ² = .16 F = 19.06 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	4.50	.13		34.24	<.001
Fluidity	.44	.07	.29	6.68	<.001
Gender	.69	.18	.24	3.93	<.001
Basic WP	-.15	.21	-.05	0.74	.46
Robust WP	.81	.18	.27	4.57	<.001
Gender x BWP	-.27	.27	-.06	1.01	.31
Gender x RWP	-.21	.25	-.05	0.81	.42
Total Score					
R ² = .18 F = 22.10 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	9.66	.29		33.19	<.001
Fluidity	1.08	.15	.32	7.43	<.001
Gender	1.89	.39	.29	4.84	<.001
Basic WP	.43	.46	.06	0.94	.35
Robust WP	1.68	.39	.25	4.31	<.001
Gender x BWP	-.98	.60	-.10	1.65	.10
Gender x RWP	-.43	.56	-.05	0.77	.44

Modal Effect by Race/Ethnicity

Of the 706 students for whom race/ethnicity information is available, 80.2% indicated that they were white while the remaining 19.8% indicated that they were non-white or of mixed race/ethnicity. As seen in Table 30, students classified as white received significantly higher Topic Development, standard English conventions, and total writing scores. In addition, students classified as white were significantly faster keyboarders and had higher computer fluidity and computer literacy scores. However, there were not statistically significant differences in the use of computers for writing between students classified as white or non-white.

Table 30: Comparison of Writing and Computer Use Measures by Race/Ethnicity

	White (566)		Non-White (140)		p
	Mean	St. Dev.	Mean	St. Dev.	
Topic Development	6.08	2.06	5.68	2.04	.04
English Standards	5.05	1.44	4.72	1.38	.02
Total Score	11.13	3.25	10.40	3.15	.02
Keyboard Speed	124.64	55.94	107.34	55.36	.002
Fluidity	.10	.02	-.15	1.16	.01
Literacy	.12	.94	-.27	1.12	.001
Writing Use	.00	.98	.03	1.10	.77

To examine the extent to which the mode of administration affected the performance of students classified as white or non-white, separate ANOVAs were performed within each race/ethnicity category. As seen in Table 31, there were no significant differences among administration modes for any of the writing scales for students classified as non-white. For students classified as white, however, there were significant differences among administration modes for all three writing scales. Specifically, white students who performed the writing test using either type of word processor performed significantly higher on the topic development scale and the total score than students who composed their response on paper.

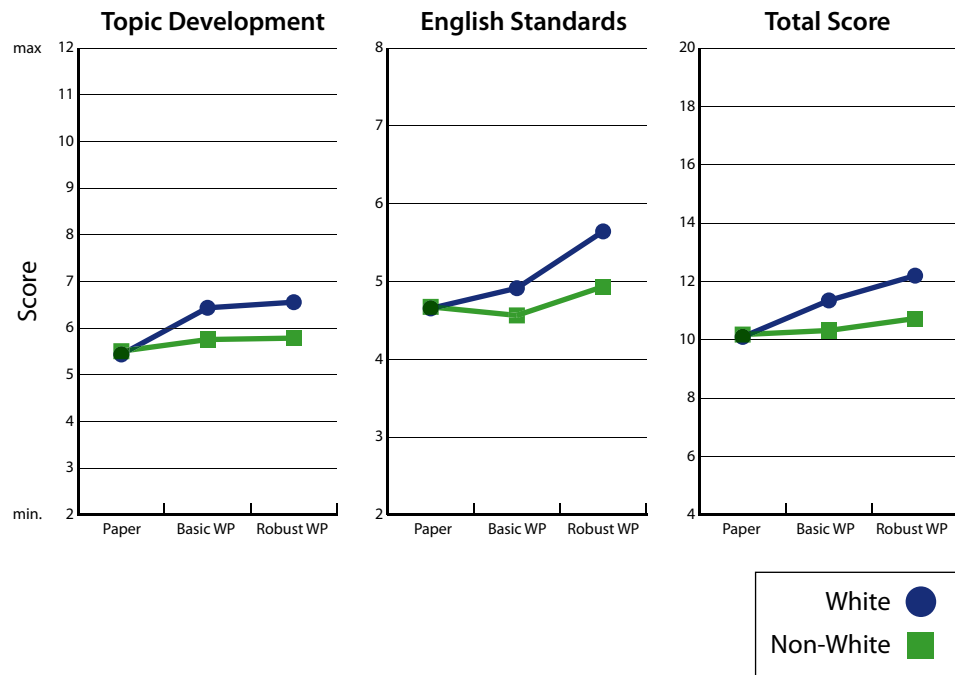
(Table 31 is shown on the following page.)

Table 31: Comparison of Mean Writing Test Score Across Presentation Mode by Race/Ethnicity

	N	All Students	Paper	Basic WP	Robust WP	F	p
Topic Development							
Non-White	140	5.68	5.50	5.75	5.78	.26	.769
White ^{1,2}	566	6.08	5.43	6.43	6.55	19.17	<.001
English Standards							
Non-White	140	4.72	4.67	4.56	4.93	.89	.413
White ^{2,3}	566	5.05	4.65	4.91	5.64	26.87	<.001
Total Score							
Non-White	140	10.40	10.17	10.31	10.72	.37	.69
White ^{1,2,3}	566	11.13	10.09	11.34	12.19	23.30	<.001

- 1 significant difference between Paper and Basic WP at p<.05 using Tukey's post-hoc mean comparison
- 2 significant difference between Paper and Robust WP at p<.05 using Tukey's post-hoc mean comparison
- 3 significant difference between Basic WP and Robust WP at p<.05 using Tukey's post-hoc mean comparison

Figure 10: Mean Writing Scores for Race/Ethnicity by Administration Mode



To examine whether the interaction effect that appears to occur between race/ethnicity and mode of administration is significant (Figure 10), a series of regression analyses was conducted in which each writing scale served as the dependent variable and computer fluidity, race/ethnicity, group membership served as the predictors. In addition, an interaction term for each word processing group was

also included in each model. The interaction term was created by multiplying race/ethnicity status (coded 1 for white) and word processor group status. Thus the interaction term indicated the effect of white students composing responses on a basic word processor or on a robust word processor. As seen in Table 32, computer fluidity is a significant predictor of scores for all three writing scales. In addition, the basic word processing condition is a significant negative predictor of the standard English convention scores, but not a significant predictor of topic development or total writing scores. Similarly, race/ethnicity and robust word processing are not significant predictors of scores for any of the writing scales. Interestingly, while the interaction between race/ethnicity and the basic word processing condition is not significant for any of the writing scores, the interaction between race/ethnicity and the robust word processing condition is significant for the standard English convention and total writing scores. For both writing scores, the interaction term indicates that students classified as white perform significantly better when using a robust word processor while students classified as non-white are largely unaffected by the use of a robust word processor (see Figure 10).

(Table 32 is shown on the following page.)

Table 32: Regression Model for Computer Fluidity, Race/Ethnicity, Group Membership and Race/Ethnicity-by-Administration Mode Predicting Writing Scores

Topic Development					
R ² = .12 F = 13.96 p < .001					
Variable	Coefficient	SE	Beta	T ratio	Sig.
Intercept	5.89	.32		18.42	<.001
Fluidity	.58	.09	.28	6.18	<.001
Race/Ethnicity	-.12	.34	-.02	0.35	.73
Basic WP	-.28	.47	-.06	0.59	.56
Robust WP	-.05	.43	-.01	.12	.90
Race/Eth. x BWP	.62	.50	.13	1.24	.22
Race/Eth. x RWP	.83	.47	.18	1.76	.08
English Standards					
R ² = .14 F = 16.37 p < .001					
Variable	Coefficient	SE	Beta	T ratio	Sig.
Intercept	4.94	.23		22.00	<.001
Fluidity	.40	.07	.27	6.13	<.001
Race/Ethnicity	-.08	.24	-.02	0.34	.73
Basic WP	-.65	.33	-.20	1.96	.05
Robust WP	-.07	.30	-.00	0.02	.98
Race/Eth. x BWP	.45	.35	.13	1.27	.20
Race/Eth. x RWP	.79	.33	.25	2.37	.02
Total Score					
R ² = .14 F = 16,47 p < .001					
Variable	Coefficient	SE	Beta	T ratio	Sig.
Intercept	10.83	.50		21.56	<.001
Fluidity	.99	.15	.30	6.68	<.001
Race/Ethnicity	-.20	.54	-.02	0.38	.71
Basic WP	-.92	.74	-.13	1.25	.21
Robust WP	-.06	.68	-.01	0.09	.93
Race/Eth. x BWP	1.07	.79	.14	1.35	.18
Race/Eth. x RWP	1.62	.74	.23	2.18	.03

Modal Effect by Multilingual Status

Of the 706 students for whom multilingual information is available, 11.3% indicated that they spoke a language other than English at home. As seen in Table 33, non-multilingual students received significantly higher Topic Development, standard English conventions, and total writing scores. In addition, non-multilingual students were significantly faster keyboarders and had higher computer fluidity and computer literacy scores. However, there were not statistically significant differences in the use of computers for writing between multilingual and non-multilingual students.

Table 33: Comparison of Writing and Computer Use Measures by Multilingual Status

	Non-Multilingual (627)		Multilingual (80)		p
	Mean	St. Dev.	Mean	St. Dev.	
Topic Development	6.08	2.07	5.41	1.97	.007
English Standards	5.05	1.44	4.51	1.40	.002
Total Score	11.13	3.26	9.93	3.06	.002
Keyboard Speed	122.95	56.77	108.18	49.04	.04
Fluidity	.11	.90	-.48	1.36	.001
Literacy	.10	.96	-.39	1.11	.001
Writing Use	-.01	.97	.11	1.20	.39

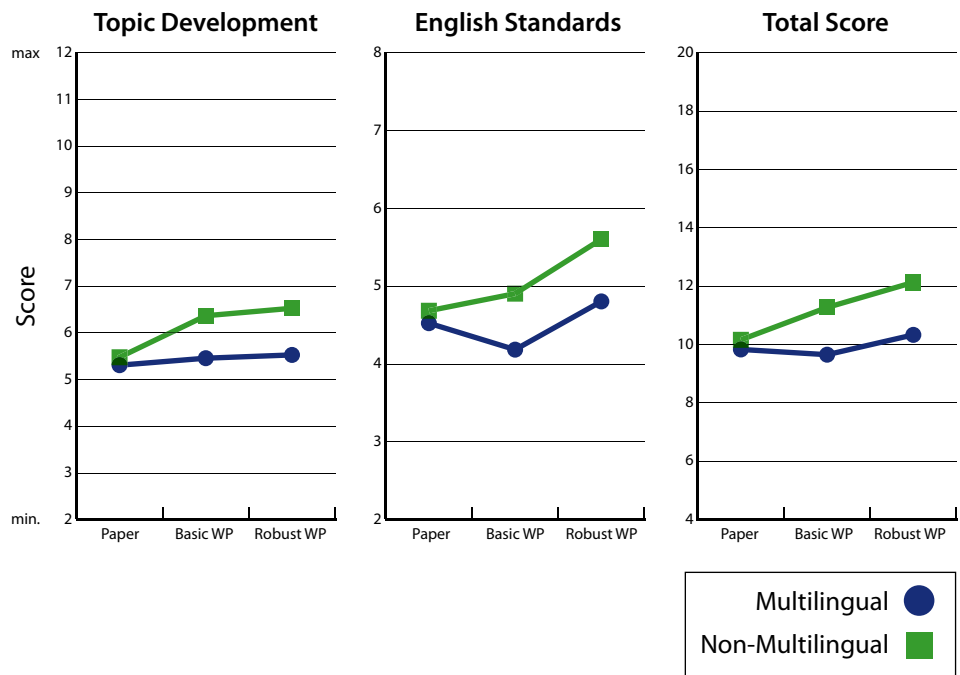
To examine the extent to which the mode of administration affected the performance of multilingual and non-multilingual students, separate ANOVAs were performed within each multilingual category. As seen in Table 34, although multilingual students received high scores when they produced essays using the robust word processor, none of the score differences were statistically significant for the three writing scales. For non-multilingual students, however, significant differences did occur across the mode of administration conditions. For all three writing scores, non-multilingual students performed significantly better using the robust word processor as compared to paper. For both the standard English conventions and total writing score, students using the robust word processor also scored significantly higher than students using the basic word processor. Finally, for topic development and total writing score, non-multilingual students using the basic word processor performed significantly better than students using paper.

Table 34: Comparison of Mean Writing Test Score Across Presentation Mode by Multilingual Status

	N	All Students	Paper	Basic WP	Robust WP	F	p
Topic Development							
Non-Multilingual ^{1,2}	627	6.08	5.47	6.36	6.52	17.52	<.001
Multilingual	80	5.41	5.30	5.45	5.52	0.09	.91
English Standards							
Non-Multilingual ^{2,3}	627	5.05	4.68	4.90	5.60	25.67	<.001
Multilingual	80	4.51	4.52	4.18	4.80	1.15	.32
Total Score							
Non-Multilingual ^{1,2,3}	627	11.13	10.15	11.26	12.12	21.72	<.001
Multilingual	80	9.93	9.82	9.64	10.32	0.32	.73

- 1 significant difference between Paper and Basic WP at p<.05 using Tukey's post-hoc mean comparison
- 2 significant difference between Paper and Robust WP at p<.05 using Tukey's post-hoc mean comparison
- 3 significant difference between Basic WP and Robust WP at p<.05 using Tukey's post-hoc mean comparison

Figure 11: Mean Writing Scores for Multilingual Status by Administration Mode



Although the graphs displayed in Figure 11 suggest that there may be an interaction between the mode of administration effect and multilingual status, the regression analyses presented in Table 35 indicate that these interactions are not statistically significant. For all three writing scales, the only significant predictors of student scores were computer fluidity and the use of a robust word processor.

Table 35: Regression Model for Computer Fluidity, Multilingual Status, Group Membership and Gender-by-Administration Mode Predicting Writing Scores

Topic Development					
R ² = .11 F = 13.01 p < .001					
Variable	Coefficient	SE	Beta	T ratio	Sig.
Intercept	5.79	.14	–	41.37	<.001
Fluidity	.57	.10	.27	6.00	<.001
Multilingual	-.01	.40	.00	0.02	.99
Basic WP	.27	.24	.06	1.15	.25
Robust WP	.71	.20	.17	3.61	<.001
Multi x BWP	-.50	.72	-.03	0.69	.49
Multi x RWP	-.66	.58	-.06	1.14	.25
English Standards					
R ² = .13 F = 15.44 p < .001					
Variable	Coefficient	SE	Beta	T ratio	Sig.
Intercept	4.89	.10	–	49.77	<.001
Fluidity	.39	.07	.26	5.82	<.001
Multilingual	-.18	.28	-.04	0.68	.50
Basic WP	-.24	.17	-.08	1.47	.14
Robust WP	.69	.14	.23	4.94	<.001
Multi x BWP	-.54	.51	-.05	1.06	.29
Multi x RWP	-.40	.40	-.05	0.98	.33
Total Score					
R ² = .13 F = 15.30 p < .001					
Variable	Coefficient	SE	Beta	T ratio	Sig.
Intercept	10.68	.22	–	48.52	<.001
Fluidity	.96	.15	.29	6.41	<.001
Multilingual	-.18	.63	-.02	0.29	.77
Basic WP	-.03	.37	.00	0.07	.94
Robust WP	1.40	.31	.21	4.50	<.001
Multi x BWP	-1.03	1.13	-.04	0.91	.36
Multi x RWP	-1.05	.91	-.06	1.17	.25

Modal Effect by IEP Status

Of the 703 students for whom IEP information is available, 13.5% indicated that they had an IEP. As seen in Table 36, non-IEP students received significantly higher Topic Development, standard English conventions, and total writing scores. In addition, non-IEP students were significantly faster keyboarders and had higher computer literacy scores. Conversely, non-IEP students reported more frequent use of a computer for writing. There were no significant differences in computer fluidity scores between IEP and non-IEP students.

Table 36: Comparison of Writing and Computer Use Measures by IEP Status

	Non-IEP (608)		IEP (95)		p
	Mean	St. Dev.	Mean	St. Dev.	
Topic Development	6.16	2.05	5.11	1.89	<.001
English Standards	5.11	1.42	4.23	1.33	<.001
Total Score	11.27	3.21	9.34	2.96	<.001
Keyboard Speed	125.34	56.27	96.60	48.74	<.001
Fluidity	.06	.97	.02	.99	.75
Literacy	.10	.96	-.35	1.07	<.001
Writing Use	-.04	.98	.26	1.10	.02

To examine the extent to which the mode of administration affected the performance of IEP and non-IEP students, separate ANOVAs were performed within each IEP category. As seen in Table 37, both IEP and non-IEP students received significantly higher scores when using a robust word processor as compared to paper. Similarly, both IEP and non-IEP students received significantly higher topic development and total writing scores when they produced their response using a basic word processor as compared to paper. In addition, non-IEP students performed significantly better on the standard English conventions and total writing score scales when they used a robust word processor as compared to a basic word processor.

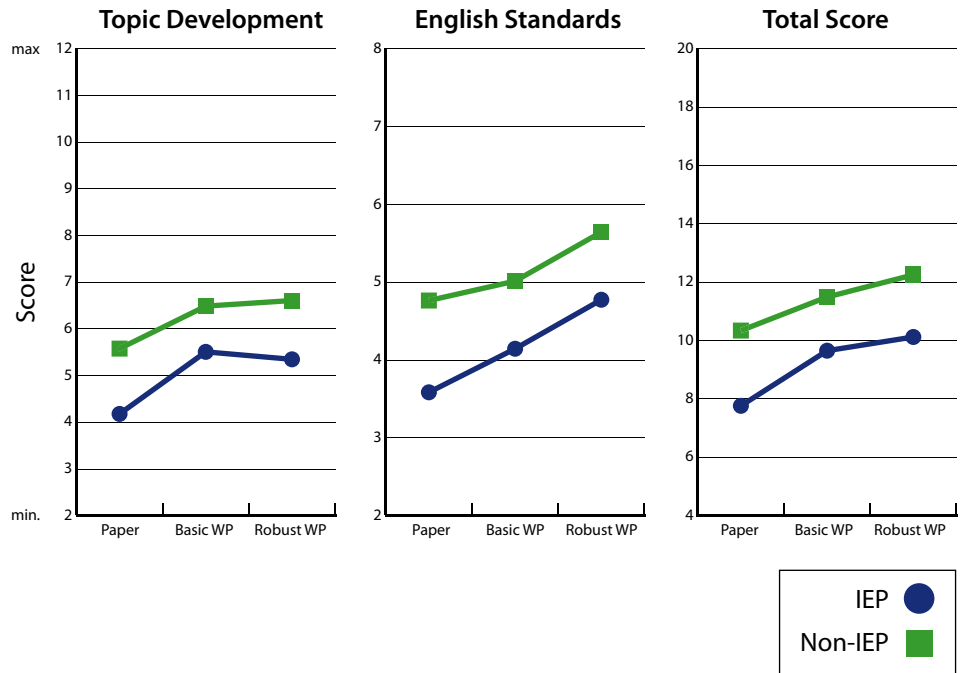
(Table 37 is shown on the following page.)

Table 37: Comparison of Mean Writing Test Score Across Presentation Mode by IEP Status

	N	All Students	Paper	Basic WP	Robust WP	F	p
Topic Development							
Non-IEP ^{1,2}	608	6.16	5.57	6.48	6.60	17.37	<.001
IEP ^{1,2}	95	5.11	4.17	5.50	5.34	4.29	.02
English Standards							
Non-IEP ^{2,3}	608	5.11	4.76	5.01	5.64	23.58	<.001
IEP ²	95	4.23	3.58	4.14	4.77	6.48	.002
Total Score							
Non-IEP ^{1,2,3}	608	11.27	10.33	11.48	12.24	21.82	<.001
IEP ^{1,2}	95	9.34	7.75	9.64	10.11	5.29	.007

- 1 significant difference between Paper and Basic WP at p<.05 using Tukey's post-hoc mean comparison
- 2 significant difference between Paper and Robust WP at p<.05 using Tukey's post-hoc mean comparison
- 3 significant difference between Basic WP and Robust WP at p<.05 using Tukey's post-hoc mean comparison

Figure 12: Mean Writing Scores for IEP Status by Administration Mode



The graphs displayed in Figure 12 suggest that there is not an interaction between the mode of administration effect and IEP status. As seen in Table 38, no statistically significant interactions were found when IEP status, computer fluidity, group membership, and the interaction between group membership and IEP status were used to predict scores for each of the writing scales. For the regression models displayed in Table 38, the interaction terms represent the effect of IEP students performing the test under either the basic or robust word processing

conditions. For all three writing scales, the interaction terms are not statistically significant. For all three models, however, computer fluidity and the robust word processing conditions are the only significant predictors of students' writing scores. In addition, IEP status had a statistically significant negative effect on scores for all three writing scales.

Table 38: Regression Model for Computer Fluidity, IEP Status, Group Membership and Gender-by-Administration Mode Predicting Writing Scores

Topic Development					
R ² = .13 F = 15.88 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	5.85	.14		42.77	<.001
Fluidity	.55	.09	.26	5.94	<.001
IEP	-1.01	.46	-.17	2.24	.03
Basic WP	.35	.24	.08	1.47	.14
Robust WP	.74	.20	.17	3.76	<.001
IEP x BWP	.27	.59	.03	0.46	.65
IEP x RWP	-.04	.58	-.00	0.06	.95
English Standards					
R ² = .16 F = 19.23 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	4.93	.10		51.59	<.001
Fluidity	.38	.07	.26	5.91	<.001
IEP	-.98	.32	-.23	3.07	.002
Basic WP	-.17	.17	-.05	1.01	.31
Robust WP	.69	.14	.23	5.05	<.001
IEP x BWP	.20	.41	.03	.49	.62
IEP x RWP	.23	.41	.04	.56	.57
Total Score					
R ² = .16 F = 19.17 p < .001					
Variable	Coefficient	SE	Beta	t ratio	Sig.
Intercept	10.77	.21		50.37	<.001
Fluidity	.93	.15	.28	6.43	<.001
IEP	-1.99	.71	-.21	-2.79	.005
Basic WP	.18	.38	.03	.49	.63
Robust WP	1.43	.31	.21	4.66	<.001
IEP x BWP	.48	.93	.03	.51	.61
IEP x RWP	.20	.91	.01	.21	.83

Discussion

A substantial body of research provides evidence that regular use of computers throughout the writing process leads to improved writing quality (Goldberg, Russell, & Cook, 2003). To provide students with increased access to computers in school, American schools continue to make substantial investments to acquire additional computers. Over the past two years, several states have even begun exploring the feasibility and effect of providing a laptop for each student. As an example, two years ago Maine launched a statewide laptop program that has provided every seventh and eighth grader in the state with a laptop. More recently, New Hampshire, Vermont, Texas, Michigan, and Massachusetts have each launched pilot studies in which students in a small number of schools are provided with a laptop.

Despite the evidence supporting instructional use of computers for writing and continued efforts to provide students sufficient access to computers in school, the majority of today's state-sponsored writing tests are administered on paper. In fact, the only exception to this practice occurred last year in Maine when students in a sample of 60 schools were allowed to compose their essay using a state provided laptop. A series of studies conducted since the mid-1990's, however, provides evidence that students who are accustomed to writing with computers under-perform when they take writing tests on paper (Russell & Haney, 1997). Conversely, students who are not accustomed to writing with a computer under-perform when forced to compose their essay using a computer (Russell, 1999; Russell & Plati, 2001; Russell & Plati, 2002).

To date, this research on the effect of mode of administration has focused on relatively small samples of students who attended school in two Massachusetts districts. The study presented in this report was conducted to both replicate previous research and to examine its generalizability across a substantially larger and more diverse sample of students. As described in greater detail above, the sample employed in this study ($n=966$) was more than three times larger than the samples employed in previous studies and included eighth grade students from rural, suburban, and urban schools located in Rhode Island, New Hampshire, and Vermont. The sample also included students with Individual Education Plans (IEP), students who spoke a language other than English at home, and a substantial percentage of minority students. The inclusion of these students allowed the mode of administration effect to be examined for a variety of sub-samples of students.

Overall, the findings reported in greater detail above are generally consistent with previous research. As seen in Table 3, on average, students performed better on the topic development component of the scoring criteria when they were able to compose responses using a basic word processor (without spell or grammar check) as compared to composing their responses using paper-and-pencil. However, there was no difference in scores between the basic word processor and paper-and-pencil on the standard English conventions portion of the scoring criteria. Students, however, tended to perform even better when they were able to compose

their essays using a more robust word processor (Microsoft Word) that allowed students to use spell and grammar check. Interestingly, the improved performance that occurred when using the robust word processor effected both the topic development and standard English conventions scores. However, the largest differential effect was found for the standard English conventions scores, which suggests that the spell and/or grammar check assisted students in correcting mechanical aspects of their writing.

In past research, a positive effect of composing essays with computers was found for students with more advanced computer skills while a negative effect was found for students with weaker computer skills. As a result, the positive and negative effects tended to cancel each other out when examining the effect across the entire sample of students. In the study presented here, the mode of administration effect was found for the entire sample. This, however, may result from students in this sample being more computer savvy than students in past samples. As an example, students included in a study conducted in 1998 had an average keyboard speed of 16.5 words per minute. That study found that the mode of administration effect tended to occur for students who could keyboard approximately 20 words per minute or more (Russell, 1999). Students in the study reported here had an average keyboard speed of 26.6 words per minute. Given that six years have passed between these two studies, during which time students have had increased access to and use of computers in school and at home, it is not surprising that their keyboarding skills have improved. Moreover, given the increased keyboarding speed of students included in this study, it is not surprising that the mode of administration effect was detected for the entire sample.

Interestingly, a differential mode of administration effect was found for students with less developed keyboarding skills, computer fluidity, and computer literacy. As reported in Tables 5, 9, and 14, significant effects were not found for students who were classified as having low keyboarding, computer fluidity, or computer literacy skills. Substantial effects, however, were found for students with medium and high keyboarding, fluidity, and literacy skills. As with the total group, the effects for the basic word processor were found for the topic development scale while the effects for the robust word processor occurred for both the topic development and standard English convention scales. It should be noted, however, that while no significant effects were found for students with lower computer skills, this pattern differs from previous studies which found significant negative effects for students with low computer skills. Again, this difference may result from students in this study having higher computer skills in general. Since the classification of students into high, medium, and low keyboarding, fluidity, and literacy groups was normative (that is relative to skills within the sample), students in the low group for this study had substantially stronger computer skills than students in the low group in previous studies. As a result, students in the low group for this study likely possessed sufficient computer skills that enabled them to compose an essay using a word processor without negatively affecting their performance.

It should be noted, however, that this pattern was less consistent when students were classified by the frequency with which they report using computers for writing, more generally in school, or for recreational purposes at home. For both the computer use for writing and computer use in school measures, significant effects were found for low, medium and high computer use for writing. For the computer use at home measure, the effects were not evident for students classified into the low use group, but were found for the medium and high use categories. Given that there was a stronger relationship between actual skills and literacy versus reported use of computers, future decisions about who might benefit by composing essay responses on a computer should focus more on actual computer-related skills than students' perceived frequency of computer use.

To examine whether the mode of administration effect interacted with background variables, a series of analyses was conducted in which the effect was estimated separately by gender, race/ethnicity, multilingual status, and IEP status. Although females tended to perform better than males, there was no interaction between gender and the mode of administration effect. Similarly, while IEP students tended to perform worse, on average, than non-IEP students, an interaction between IEP status and the mode of administration effect was not detected. While the graphical display for multilingual students suggest that an interaction effect may have occurred (see Figure 11), the regression analysis indicate that this effect was not statistically significant. It should be noted, however, that the sample size for multilingual students was relatively small which may have impacted the statistical power of the interaction analysis. Finally, a significant interaction effect was found for race/ethnicity. Specifically, while students classified as white scored higher, on average, than students classified as non-white, the interaction effect indicated that white students tended to benefit substantially more when composing their essay using the robust word processor as compared to non-white students. It should also be noted that, despite significant differences in the computer skills of white and non-white students, the interaction effect remained significant after controlling for differences in computer fluidity.

Based on the findings of this study and findings of previous research on this topic, it is apparent that writing tests administered on paper provide different information about student writing skills as compared to the same tests administered on computer. Given that students are increasingly using computers to compose written text in and out school, it seems appropriate to begin transitioning writing tests to a computer format. Whereas previous research on this topic suggested that a transition to computer-based writing tests might penalize students with low computer skills, the study presented here suggests that the vast majority of students have acquired sufficient computer skills such that even students with relatively low levels of computer skills generally do not underperform when composing an essay on paper. While there are undoubtedly some students in the general population who would be placed at a disadvantage by a computer-based writing test, this percentage of students is grossly outweighed by the substantial

percentage of students in this study who are disadvantaged by a paper-based writing test. That is, if the sample of students included in this study is representative of students in New England, at a minimum, then it appears that the majority of students have obtained sufficient computer skills to benefit or at least not be harmed by composing an essay on computer.

One concern that remains, however, is the interaction effect that was detected for students who were classified as non-white and that was visible (although not statistically significant) for multilingual students. Again, while it is unclear why this interaction may be occurring, its presence warrants further investigation. Yet, it should be noted that despite this interaction, neither students classified as non-white nor multilingual students were disadvantaged by composing their essays using either a basic or robust word processor. Instead, the interaction occurred because students classified as white or non-multilingual performed noticeably better when using a basic word processor and significantly better when using a robust word processor.

Finally, as test administrators contemplate transitioning writing tests to a computer format, it will be important to consider what is intended to be measured through the standard English conventions portion of the scoring criteria. If the primary aim of this scoring dimension is to identify the extent to which students can correctly spell and form grammatically correct sentences without external assistance, then it will be important to deny students access to spell and grammar check. However, if this dimension is intended to measure students' ability to use modern tools, including paper-based dictionaries and computer-based spell and grammar checks, and to compose essays with minimal spelling and grammatical errors, then the use of robust word processors may be appropriate.

In conclusion, this study provides further evidence that computer savvy students with moderate keyboarding, computer fluidity, and computer literacy skills produce higher quality essays when they are allowed to use a word processor during a writing test. This study, however, also provides evidence that this mode of administration effect may differ for students who are classified as non-white and/or who are multilingual. Finally, this study provides evidence that the mode of administration effect is larger when students have access to a robust word processor that includes spell and grammar check. Collectively, these findings suggest that state testing programs should consider transitioning their writing tests to a computer format. However, in doing so, serious consideration should be given to the construct being measured by the standard English convention portion of scoring rubrics and the appropriateness of providing students access to spell and grammar checks. In addition, further research should be conducted to develop a better understanding as to why the interaction between the mode of administration effect and race/ethnicity and, perhaps, multilingual status may be occurring.



References

- Bangert-Downs, R.L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research*, 63(1), 69–93.
- Becker, H. J. (1999). *Internet Use by Teachers: Conditions of Professional Use and Teacher-Directed Student Use*. Irvine, CA: Center for Research on Information Technology and Organizations.
- Bunderson, C., Inouye, D., & Olsen, J. (1989). The four generations of computerized educational measurement. In Linn, R.L., *Educational Measurement* (3rd ed.), Washington, D.C.: American Council on Education, pp. 367–407.
- Choi, S.W., & Tinkler, T. (April, 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.
- Clariana, R. & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, 33(5).
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Daiute, C. (1986). Physical and cognitive factors in revising: insights from studies with computers. *Research in the Teaching of English*, 20 (May), p. 141–59.
- Etchison, C. (1989). Word processing: A helpful tool for basic writers. *Computers and Composition*, 6(2), 33–43.
- Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992–2002. *Journal of Technology, Learning, and Assessment*, 2(1). Available from <http://www.jtla.org>
- Haas, C. & Hayes, J. (1986). What did I just say? Reading problems in writing with the machine. *Research in the Teaching of English*, 20(1), 22–35.
- Haertel, E. & Wiley, D. (2003, August). *Comparability issues when scores are produced under varying testing conditions*. Paper presented at the Psychometric Conference on Validity and Accommodations in College Park, MD.
- Haney, W., Madaus, G., & Lyons, R. (1993). *The Fractured Marketplace for Standardized Testing*. Boston, MA: Kluwer Academic Publishers.
- Hannafin, M. J. & Dalton, D. W. (1987). The effects of word processing on written composition. *The Journal of Educational Research*, 80 (July/Aug.) p. 338–342.

- Kerchner, L. B. & Kistingner, B. J. (1984). Language processing/word processing: Written expression, computers, and learning disabled students. *Learning Disability Quarterly*, 7, 329–335.
- Mazzeo, J. & Harvey, A.L. (1988). *The equivalence of scores from automated and conventional educational and psychological tests: A review of the literature* (College Board Rep. No. 88-8).
- Mead, A. D & Drasgow, F. (1993) Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114(3), 449–458.
- Pommerich, M. & Burden, T. (2000, April). *From simulation to Application: Examinees react to computerized testing*. Paper presented at the National Council on Measurement in Education conference in New Orleans. Retrieved on August 6, 2003, from <http://home.att.net/~pommie/NCME00.pdf>
- Pommerich, M. (2004). Developing computerized versions of paper-and-pencil tests: Some mode effects for passage-based tests. *Journal of Technology, Learning, and Assessment*, 2(6). Available from <http://www.jtla.org>
- Russell, M. & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-Pencil. *Educational Policy Analysis Archives*, 5(3). Available: <http://epaa.asu.edu/epaa/v5n3.html>
- Russell, M. & Plati, T. (2000). *Mode of Administration Effects on MCAS Composition Performance for Grades Four, Eight, and Ten*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy. Available: <http://www.bc.edu/research/nbetpp/reports.html>.
- Russell, M. & Plati, T. (2001). Effects of computer versus paper administration of a state-mandated writing assessment. *Teachers College Record*. Available: <http://www.tcrecord.org/Content.asp?ContentID=10709>
- Russell, M. & Plati, T. (2002). Does it matter with which I write: Comparing performance on paper, computer and portable writing devices. *Current Issues in Education [On-line]*, 5(4). Available: <http://cie.ed.asu.edu/volume5/number4/>
- Russell, M. (1999). Testing Writing on Computers: A Follow-up Study Comparing Performance on Computer and on Paper. *Educational Policy Analysis Archives*, 7(20).
- Russell, M., Bebell, D., & O'Dwyer, L. (2003). *Use, support, and effect of instructional technology study: An overview of the USEIT study and the participating districts*. Boston, MA: Technology and Assessment Study Collaborative. Available for download at http://www.intasc.org/PDF/useit_r1.pdf
- The Editors (2003). Tech's Answer to Testing. *Education Week*. XXII(35), 8–10.

Vacc, N. N. (1987). Word processor versus handwriting: A comparative study of writing samples produced by mildly mentally handicapped students. *Exceptional Children*, 54(2), 156–165.

Williamson, M. L. & Pence, P. (1989). Word processing and student writers. In B. K. Briten & S. M. Glynn (Eds.), *Computer Writing Environments: Theory, Research, and Design* (pp. 96–127). Hillsdale, NJ: Lawrence Erlbaum & Associates.

