Does it Matter with What I Write?
Comparing Performance on Paper, Computer and Portable Writing Devices

Michael Russell
Boston College

Tom Plati
Wellesley Public Schools

This study builds on three previous studies (Russell, 1999; Russell & Haney, 1997; Russell & Plati, 2001) to examine the effect of administering extended composition test items on paper, on computer, or on a portable writing device has on student performance. This study employs writing items from the 1999 Massachusetts Comprehensive Assessment System (MCAS) to examine the mode of administration effect in grades four and eight. Similar to previous studies, this article finds that open-ended Language Arts items that require students to generate responses using paper and pencil, severely underestimate the achievement of fourth grade students accustomed to writing using a computer. This study also finds that open-ended tests administered on paper underestimate the achievement of eighth grade students accustomed to writing with an eMate (a portable writing device). Combining the effects found in this study with those found in Russell's 1999 study, this article estimates that the MCAS Language Arts test underestimates the performance of students accustomed to writing using a computer by four to eight points on an eighty point scale. This article concludes by recommending that state testing programs that employ open-ended items in Language Arts provide students with the option of composing responses using the writing tools with which they are accustomed to working.

Over the past decade, the presence of computers in schools has increased exponentially. One of the many areas in which teachers use computers is writing. As several studies demonstrate, regular use of computers for writing over extended periods of time can lead to significant improvements in students' writing skills (see Russell & Plati, 2001 for a more detailed review of the literature on computers and writing). To capitalize on these benefits, a few schools have made laptop computers available to all of their students. In most schools, however, the relatively high cost of desktop and laptop computers prohibit schools from acquiring a sufficient quantity of computers for all students to use simultaneously.

To provide an entire class of students access to computers, some schools place large numbers of computers into a computer lab. While this strategy succeeds in providing large numbers of students simultaneous access to computers, it may encourage teachers to treat writing with computers as a special event rather than central to their classroom teaching

(Kleiman, 2000). However, the recent introduction of portable writing devices, such as AlphaSmarts, DreamWriters and eMates, provide schools with a more affordable option that allows all students to write with a word processor in their classrooms.

Schools are able to purchase about eight portable writing devices for the same price as one desktop computer. Instead of sharing a limited number of computers in a classroom or taking turns using computers in a lab, portable writing devices allow all students within a classroom to write using a word processor. Basic word processors allow students to compose, edit, cut-copy-and-paste text, print, and in some cases perform spell-checking.

Although portable writing devices are relatively new to schools, their presence is increasing rapidly. As Table 1 displays, sales of Dreamwriters more than doubled each year between 1995 and 1998. Similarly, the presence of AlphaSmarts in schools has increased steadily over the past three years such that there are now just over 10,000 AlphaSmarts in Massachusetts schools alone. Although the number of computers in schools remains greater than the quantity of portable writing devices, schools are rapidly turning to portable writing devices as a strategy for providing all students regular and extended time writing with a word processor.

Table 1:
Sales growth of DreamWriter and related peripherals

Year	Sales in Canadian $ Millions
1994	1.12
1995	3.85
1996	9.50
1997	20.58
1998	34.03

(NTS Computer Systems, 1998)

During the past decade, the prominence of educational testing has also increased. According to Education Week, all 50 states now use statewide tests to assess student performance in different subject areas (Meyer, Orlofsky, Skinner & Spicer, 2002). As a report issued by the Council of Chief State School Officers (1998, p. 17) stated, "Possibly the greatest changes in the nature of state student assessment programs have taken place in the 1990s as more states have incorporated open-ended and performance exercises into their tests, and moved away from reliance on only multiple-choice tests." Currently, 46 state testing programs include sections in which students must write extended answers or written explanations of their work (Meyer, Orlofsky, Skinner & Spicer, 2002).

Recent research, however, demonstrates that these two strategies for improving education, namely state level testing programs and writing on computers, may work against each other. Although several studies have concluded that multiple-choice tests administered on paper or on computer yield equivalent scores (Bunderson, Inouye & Olsen, 1989; Mead & Drasgow, 1993), more recent research shows that open-ended tests administered on paper significantly under-estimate the achievement of students accustomed to writing on computers (Russell, 1999; Russell & Haney, 1997). In both studies, the effect sizes for students accustomed to writing on computer ranged from .57 to 1.25. Effect sizes of this magnitude are unusually large and of sufficient size to be of not just statistical, but also practical significance (Cohen, 1988; Wolf, 1986). Effect sizes of this magnitude, for example, imply that the score for the average student in the experimental group tested on computer exceeds that of 72 to 89 percent of the students in the control group tested via paper and pencil.

Although prior research on computer use and performance on open-ended test items administered on paper does not call into question the value of state level accountability systems, it does suggest that we should begin thinking about alternative ways of administering open-ended items. As state level accountability tests begin to transition from paper administration to computer administration, several issues arise. First, until all students are accustomed to writing on computers, a better understanding of the extent to which the mode of administration affects student performance at different grade levels must be developed. Second, given the many computing devices available, the effect of performing open-ended items on desktop computers needs to be contrasted with performance on cheaper and more portable writing devices such as eMates and AlphaSmarts. Third, before testing programs offer students the option of performing tests on paper or on computer, the extent to which handwritten versus computer printed responses influence raters' scores needs to be explored. Fourth, administrative procedures for administering tests on computers in schools must be developed.

The two studies presented here build on the previous work of Russell (1999) and Russell and Haney (1997) and address the first two issues presented above. Whereas previous studies have focused only on middle school students, the two experiments presented here focus on students in fourth and eighth grade. More importantly, whereas the previous studies have compared performance on

paper and on computer, the studies presented here examine a third mode of administration, namely portable writing devices. Finally, whereas the previous studies examined the effect on relatively short open-ended items that ranged from two to thirty minutes in length, the two studies presented here focus on extended composition items to be completed during two 45 to 60 minute blocks of time.

More specifically, the first experiment presented here focuses on the mode of administration effect for eMates in grade eight. The second study presented here examines the mode of administration effect for both desktop computers and AlphaSmarts in grade four. Three future articles will: a) examine the mode of administration effect for special education students; b) examine the influence computer versus handwritten responses have on raters' scores; and c) explore policy and administration procedures that may reduce the effect mode of administration has on students' performance.

Study Design

To explore the mode of administration effect for portable writing devices, two separate experiments were conducted. The first experiment occurred with students in grade eight and compared performance on paper with performance on an eMate. The second experiment was conducted with fourth grade students and compared students' performance on paper, on a desktop computer, and on an AlphaSmart.

An AlphaSmart is a portable word processing device that allows students to enter text into a small window that displays four lines of text with forty characters per line. Students may edit text on the AlphaSmart using arrow keys and the delete button. Cutting and pasting are not available to students using an AlphaSmart. To better enable students to revise their composition, students who composed their rough drafts on an AlphaSmart were allowed to edit and finalize their composition on a desktop computer.

An eMate is also a portable word processor, but differs from an AlphaSmart in three important ways. First, the screen is capable of displaying up to twelve lines of text with sixty characters per line. Second, students may use a stylus to select blocks of text and to place the cursor in different locations. Third, in addition to allowing students to cut, copy and paste, eMates also provide a basic spell-checker.

Students in both grade levels responded to an extended composition item from the 1999 Massachusetts Comprehensive Assessment System (MCAS). In both studies, three types of background information were collected for all students: prior grades in English, prior computer use, and keyboarding speed.

The study occurred in three stages. During stage 1, prior English grades were collected for each student. For all students, year-end grades from the previous year and mid-term grades from the current year were collected. The course-end grades were used during the stratified group assignment process and the mid-term grades were used as covariates during some analyses. During stage 2, all students completed the computer use survey and performed the keyboarding test. During stage 3, students performed the composition item.

To the extent possible, the same administration procedures were employed in this study as occurred during the 1999 MCAS administration. In the actual MCAS composition administration, students completed a composition item during two sessions. During the first session, students composed a first draft. After a fifteen-minute break, students then revised and finalized their composition during a second writing session. Both sessions were designed to last for forty-five minutes, but students were given additional time as needed. In some cases, students were reported to take up to an additional hour to complete the composition item. In this study, time constraints and scheduling conflicts challenged efforts to replicate the MCAS composition administration. In grade eight, only two hours of total testing time were available. For this reason, the two sessions were completed during two consecutive fifty-minute blocks. In grade four, students working on paper and on computer completed the two sessions on the same day. Students who composed their first draft on AlphaSmarts were not able to access a computer to edit their final drafts until the second day.

Sampling and Group Assignment

All students included in this study attended Wellesley Public Schools, a suburban school district located outside of Boston. Within the district, half of the eighth grade students attending the Middle School participated in the study (the second half of the class participated in a related study that compared performance on a desktop computer with performance on paper). In fourth grade, students attending three of the six elementary schools participated.

Within each grade level, the process of assigning students to groups was identical. Students' prior grade in English was used to stratify participating students within each grade level. Students within each stratum were then randomly assigned to groups. In grade eight, the composition item was administered in two formats, namely on paper and on eMate. Thus, two groups were formed in grade eight. In grade four, the composition item was administered in three formats: on paper, on a

desktop computer, and on an AlphaSmart. Thus, three groups were formed in grade four.

Table 2 summarizes the study conducted within each grade level and indicates the number of students assigned to each group.

Table 2:
Summary of Study Designs

	Paper	Computer	AlphaSmart	eMate
Grade 4	49	50	53	--
Grade 8	42	--	--	42

Prior Computer Use

In addition to performing the composition item, all students completed a computer use questionnaire and a keyboarding test. The computer use questionnaire focused on students' use of computers at home, in school, and during their normal writing process. In addition, the questionnaire collected information about students' use of eMates or AlphaSmarts in school and during the writing process. Finally, the questionnaire queried students about their preference for taking a writing test on: a) paper or computer, and b) paper or eMate/AlphaSmart.

Keyboarding Test

To measure keyboarding skills, all students performed a computer based keyboarding test. The keyboarding test contained two passages. Students had two minutes to type each passage verbatim into the computer. Words per minute unadjusted for errors were averaged across the two passages and were used to estimate students' keyboarding speed. For the grade eight students, both keyboarding passages were taken directly from encyclopedia articles to assure that the reading level was not too difficult. For the grade four students, the keyboarding passages were taken from a book read in many fourth grade classes. Although there is considerable debate about how to quantify keyboarding ability (see West, 1968, 1983; Russon & Wanous, 1973; Arnold, et al, 1997; and Robinson, et al, 1979), for the purposes of this study, students' average words per minute (WPM) was recorded uncorrected for errors.

Scoring

All responses were scored independently by two raters. Of the nine raters employed for this study, seven were full time classroom teachers and two were advanced doctoral students in an educational research program. All of the raters were blind to the study design, student identities, and the mode on which student responses were created. All raters participated in a one-and-one-half to two hour training session prior to scoring student responses.

For all of the items, the scoring criteria developed for MCAS were used. The MCAS scoring guidelines for the composition items focused on two areas of writing, namely Topic/Idea Development and Standard English Conventions. The scale for Topic Development ranged from 1 to 6 and the scale for English Conventions ranged from 1 to 4, with one representing the lowest level of performance for both scales. Table 3 presents the category descriptions for each point on the two scales.

Table 3:
Category Descriptions for MCAS Composition Rubrics

Score	Topic Development	English Standards
1	Little topic/idea development, organization, and/or details Little or no awareness of audience and/or task	Errors seriously interfere with communication AND Little control of sentence structure, grammar and usage and mechanics
2	Limited or weak topic/idea development, organization, and/or details Limited awareness of audience and/or task	Errors interfere somewhat with communication and/or Too many errors relative to the length of the essay or complexity of sentence structure, grammar and usage, and mechanics
3	Rudimentary topic/idea development and/or organization Basic supporting details Simplistic language	Errors do not interfere with communication and/or Few errors relative to length of essay or complexity of sentence structure, grammar and usage, and mechanics
4	Moderate topic/idea development and organization Adequate, relevant details Some variety in language	Control of sentence structure, grammar and usage, and mechanics (length and complexity of essay provide opportunity for students to show control of standard English conventions)
5	Full topic/idea development Logical organization Strong details Appropriate use of language	
6	Rich topic/idea development Careful and/or subtle organization Effective/rich use of language	

(Massachusetts Department of Education, 1999a)

In addition to the general descriptions, MCAS also provides anchor papers and benchmark papers for each category. These exemplars are grade

level specific and respond to the prompt administered at each grade level.

To reduce the influence handwriting has on raters' scores (Powers, Fowles, Farnum & Ramsey, 1994), all responses to the open-ended items administered on paper were transcribed verbatim into computer text. The transcribed responses were randomly intermixed with the computer responses. All student responses were formatted with the same font, font size, line spacing and line width. In this way, the influence the mode of response might have on the scoring process was eliminated.

Scoring guidelines designed for each item were used to score student responses. All responses were scored by two raters. To reduce the influence any one pair of raters might have on student scores, student responses were spiraled such that each rater was paired with all other raters and all raters scored an equal number of responses produced in each mode. At the conclusion of the scoring process, scores awarded by two raters were added together to produce a Topic Development scale that ranged from two to twelve and an English Standards scale that ranged from two to eight.

To estimate inter-rater reliability, the original scores from both raters were used. The resulting scores were compared both via correlation and percent agreement methods. Table 4 shows that for most items the correlation between the two raters' scores ranged from .57 to .68. Agreement within one point ranged from 88 % to 100%. Although the inter-rater correlations were lower than desired, they suggest that when discrepancies arose, one set of raters was not consistently more or less lenient than the second set of raters. Although no information has been published regarding inter-rater reliability of composition scores for the actual administration of the MCAS composition items, the extent to which raters were within one point of agreement is similar to the frequency of agreement obtained for the actual MCAS short-answer, open-ended items (Massachusetts Department of Education, 1999b).

Table 4:
Inter-rater Reliability for Open-Ended Items

	Correlation	% Within 1 Point
Grade 4		
Topic Development	.57	88%
English Standards	.68	96%
Grade 8		
Topic Development	.67	100%
English Standards	.59	94%

Results

The first of these two studies explores the relationships between prior computer use and performance on an extended composition item administered on paper and on a portable writing device in eighth grade. The second study is similar to the first except that it focuses on fourth grade students and includes a third mode of administration, namely desktop computers. To examine these relationships, three types of analyses were performed within each grade level. First, to compare performance among groups, independent sample t-tests were employed in grade eight and analyses of variance (ANOVA) were employed in grade four. Second, total group regression analyses were performed to estimate the mode of administration effect controlling for differences in prior achievement. Third, sub-group regression analyses were performed to examine the group effect at different levels of keyboarding speed. Before the results of these analyses are described, summary statistics are presented.

Summary Statistics

Summary statistics are presented for each grade level included in this study. For the student questionnaire, keyboarding test, and English grades, summary statistics are based on all students included within each grade level.

Keyboarding test. The keyboarding test contained two passages. As described above, the total number of words typed for each passage was calculated and divided by 4 to yield the number of words typed per minute (WPM) for each student. Due to the passage length, the maximum keyboarding speed students in grade eight could obtain was 59 words per minute. Table 5 indicates that the mean WPM was approximately 24 in grade four to 29 in grade eight.

Table 5:
Summary Statistics for the Keyboarding Test

	N	Mean WPM	Std Dev	Min	Max
Grade 4	152	23.71	9.91	5	62
Grade 8	84	28.95	8.59	10	59

Student questionnaire. The student questionnaire contained 12 questions. Although comparative data is not available, Tables 6a and 6b suggest that on average, students in both grade levels included in this study had substantial experience working with computers. The vast majority of students reported using a computer for three or more years, using computers in schools regularly, and using a computer in their home nearly every day.

Table 6a:

Summary Statistics for the Student Questionnaire - Grade 4

	Never	Less than one	One	Two	Three	Four or more
Years using computer	1%	3%	9%	26%	74%	

	Never	Less than 1 hour/week	1-2 hours/week	2-5 hours/week	1-2 hours/day	More than 2 hours/day
Use computer in school	3%	46%	32%	16%	2%	1%
Use computer at home	1%	18%	23%	31%	16%	11%
Use eMate/AlphaSmart in School	6%	62%	22%	0%	10%	1%

	Never	1-2 times/year	3-4 times/year	Once a month	2-3 times/month	About once a week
Compose First Draft w/ Computer	23%	17%	18%	16%	14%	12%
Edit w/ Computer	12%	14%	26%	16%	23%	9%
Type Final Draft w/ Computer	2%	11%	21%	26%	18%	23%
Compose First Draft w/ eMate/AlphaSmart	20%	20%	18%	16%	20%	7%
Edit w/ eMate/AlphaSmart	31%	26%	13%	12%	10%	8%
Type Final Draft w/ eMate/AlphaSmart	29%	29%	13%	14%	10%	5%

	Paper	Computer /eMate
Paper or Computer Preference	42%	58%
Paper or eMate/AlphaSmart Preference	57%	43%

Table 6b:

Summary Statistics for the Student Questionnaire - Grade 8

	Never	Less than one	One	Two	Three	Four or more
Years using computer	0%	0%	2%	5%	11%	82%

	Never	Less than 1 hour/week	1-2 hours/week	2-5 hours/week	1-2 hours/day	More than 2 hours/day
Use computer in school	7%	59%	28%	5%	1%	0%
Use computer at home	1%	6%	8%	28%	40%	17%
Use eMate/AlphaSmart in School	88%	12%				

	Never	1-2 times/year	3-4 times/year	Once a month	2-3 times/month	About once a week
Compose First Draft w/ Computer	8%	8%	17%	16%	19%	31%
Edit w/ Computer	15%	4%	16%	16%	21%	29%
Type Final Draft w/ Computer	2%	0%	7%	11%	29%	51%
Compose First Draft w/ eMate/AlphaSmart	10%	17%	30%	27%	13%	4%
Edit w/ eMate/AlphaSmart	11%	18%	27%	27%	13%	5%
Type Final Draft w/ eMate/AlphaSmart	5%	17%	27%	23%	21%	8%

	Paper	Computer /eMate
Paper or Computer Preference	13%	87%
Paper or eMate/AlphaSmart Preference	27%	74%

Furthermore, most students reported that they use a computer at least once a month when writing a first draft. Slightly more students reported using a computer two to three times a month to edit the first draft. And most students reported using a computer regularly to write their final draft. Similarly, most students indicated that if given the choice, they would prefer to write a paper on computer than on paper.

Table 6a and 6b also show that many students in both grade levels use portable writing devices to write first drafts often.

Indicator of prior achievement. Mid-year English grades were collected for all students included in this study. In fourth grade, students' English grades are categorized into four category scores that range from one to four. To calculate student's English grade, the scores from these four categories were totaled. The resulting scale ranged from four to sixteen.

For grade eight, alphabetic grades (e.g., A, B-, C+) were awarded. These alphabetic grades were converted to a numeric scale as indicated in Table 7.

Table 7:

Letter Grade to Numeric Grade Conversion Chart

Letter Grade	Number Grade
A+	97
A	95
A-	92
B+	87
B	85
B-	82
C+	77
C	75
C-	72
D+	67
D	65
D-	62
F	50

Table 8:

Summary Statistics for Mid-Year Grades

	N	Mean	Std Dev	Min	Max
Grade 4	152	11.0	1.91	8	16
Grade 8	84	88.8	5.73	75	97

Table 8 displays the mean and standard deviation for mid-year grades for each grade level.

Table 9:
Summary Statistics for Composition Scores

	N	Scale	Mean	Std Dev	Mean on MCAS
Grade 4					
Topic Development	152	2-12	7.64	2.04	6.74
Standard English	152	2-8	6.02	1.27	5.36
Grade 8					
Topic Development	84	2-12	8.29	1.83	7.18
Standard English	84	2-8	6.33	1.05	5.67

Comparing Performance by Mode of Administration

As is explained in greater detail above, the study designs differed for the eighth and the fourth grade. For this reason, results are reported separately for each grade level. In grade eight, independent t-tests (assuming equal variances for the two samples and using a pooled variance estimate) were employed to examine differences between the pre-assigned modes of administration. The null hypothesis was that the mean performance of the eMate and paper groups did not differ. Within grade four, an ANOVA was performed to compare mean performance for each mode. The null hypothesis for this test was that the mean performance of the paper, the computer and the AlphaSmart groups did not differ. In both grade levels, the analyses test whether performance on a portable writing device and on computer (in grade four) had a statistically significant effect on students' test scores.

To examine whether prior achievement or keyboarding skills differed between the groups of students who performed each test, independent samples t-tests were also performed for students' mid-term grades and WPM in grade eight and ANOVAs were performed in grade four. In addition, independent samples t-tests were performed to examine differences in the length of students' responses in grade eight and ANOVAs were performed in grade four. Finally, in grade eight, the amount of time students spent working on their compositions was also collected. Given concerns that the MCAS tests consume a large amount of time,

testing time was recorded to examine whether drafting and revising on eMate might reduce testing time without jeopardizing student performance. For this reason, testing time was also compared.

Grade 8. Table 10 displays results for the paper versus eMate experiment. Although the mid-term grades were slightly higher for students in the paper group, this difference was not statistically significant. On average, however, students writing with an eMate produced passages that were twenty percent longer than those composed on paper. Students who composed with an eMate also received higher scores for both Topic Development and English Standards. On average, students composing on eMate scored 1.4 points higher on Topic Development and .7 points higher on English Standards. When the two sub-scores are combined, the eMate group performed over two points higher than the paper group (see Figure 1). In addition, students writing with an eMate finished more than twenty-five minutes faster than did students writing on paper.

Table 10:

Between Group Comparisons for Paper versus eMate Experiment

	Mean	Std. Dev.	Std. Error	T-statistic	Significance	Effect Size
Mid-Term Grade						
Paper	89.0	5.7	.87			
eMate	88.5	5.9	.90	.34	.73	-.06
Topic Development						
Paper	7.6	1.5	.23			
eMate	9.0	1.9	.30	3.56	.001	.89
English Standards						
Paper	6.0	.99	.15			
eMate	6.7	.97	.15	3.57	.001	.77
Total Score						
Paper	13.6	2.29	.35			
eMate	15.7	2.70	.42	3.84	<.001	.91
Passage Length						
Paper	448	106	16.4			
eMate	536	160	24.6	2.97	.004	.82
Finish Time						
Paper	107.0	9.4	1.44			
eMate	75.8	9.1	1.41	15.42	<.001	-3.32
WPM						
Paper	29.2	8.2	1.27			
eMate	28.7	9.0	1.40	.28	.78	-.06

N for Paper Group=42
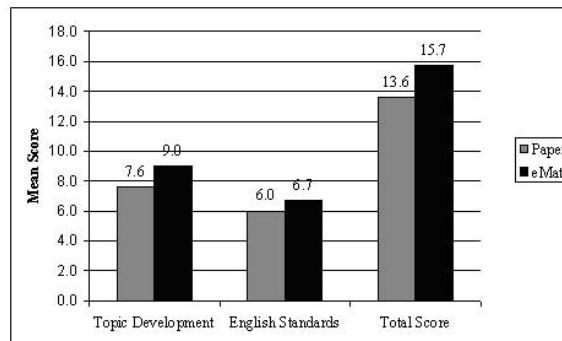N for eMate Group=42

Figure 1. Mode of administration effect on MCAS Language Arts scores - grade 8.

Note that statistical significance for the t-tests reported above was not adjusted to account for multiple comparisons. Given that seven comparisons were made within each experiment, there is an increased probability that reported differences occurred by chance. Employing the Dunn approach to multiple comparisons (see Glass & Hopkins, 1984), α for c multiple comparisons, α_{pc}, is related to simple α for a single comparison as follows:

$$\alpha_{pc} = 1 - (1 - \alpha)^{1/c}$$

Hence, for seven comparisons the adjusted value of a simple 0.05 alpha level becomes 0.007. Analogously, a simple alpha level of 0.01 for a simple comparison becomes 0.001.

Once the level of significance is adjusted for multiple comparisons, the difference in passage length, finish time, and all categories of composition scores remain statistically significant. Moreover, as shown in Table 10, these differences in composition scores represent effect sizes of .77 to .91 (Glass's delta effect size was employed). These effect sizes fall in between those reported by Russell and Haney (1997) and by Russell (1999). The effect size for the total score suggests that while about half of the students writing with an eMate scored above 15.7, less than 19% of students performing the test on paper scored above 15.7.

To control for differences in prior achievement, a multiple regression was performed. Table 11 presents the results of each test score regressed on mid-term grades and group membership. For these regression analyses, the regression coefficient (B) for group membership indicates the effect group membership has on students' performance when the effect of mid-term grade is controlled. Group membership was coded 0 for the paper group and 1 for the eMate group. A positive regression coefficient indicates that performing the test on eMate had a positive effect on students' test performance. A negative regression coefficient suggests that on average students who performed the

test on eMate scored lower than students who performed the test on paper.

Table 11 indicates that mid-term grades were a significant predictor of students' scores. For every increase of one standard score unit in mid-term grade, on average students experienced between a .50 and .54 standard score increase in their test score. Table 11 also indicates that after controlling for differences in mid-term grades, performing the composition item on an eMate had a positive impact on student scores. This impact ranges from a .39 to .41 standard score increase in student test scores. All of these effects are statistically significant.

Table 11:

Composition Scores Regression Analyses for Paper versus eMate Experiment

Topic Development

R=.62	B	Std. Error	Beta	T	Sig.
Constant	-6.7	2.49			
Mid-Year Grade	0.16	0.03	.50	5.75	<.001
Group	1.40	0.31	.39	4.41	<.001

English Standards

R=.63	B	Std. Error	Beta	T	Sig.
Constant	-2.31	1.14			
Mid-Year Grade	0.09	0.02	.51	5.88	<.001
Group	0.80	0.18	.39	4.54	<.001

Total Score

R=.66	B	Std. Error	Beta	T	Sig.
Constant	-8.98	3.50			
Mid-Year Grade	0.25	0.04	.54	6.46	<.001
Group	2.20	0.45	.41	4.93	<.001

Grade 4. The study in grade four compared performance on paper with performance on computer and on AlphaSmarts. Students were randomly assigned to one of these three groups. Table 12 indicates that mean scores for Topic Development,

Total Score and Passage Length differed among the three groups (see also Figure 2).

Table 12:

Summary Statistics by Mode of Administration

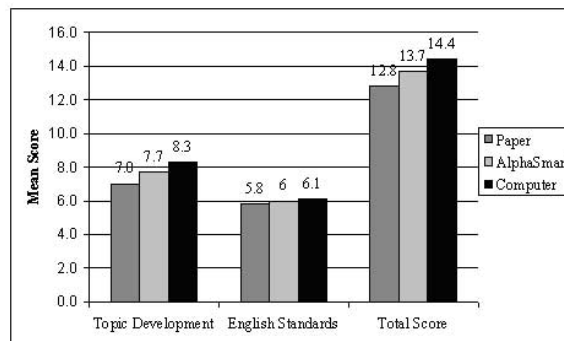|	Paper		Computer		AlphaSmart	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Mid-term Grade	10.7	1.71	11.0	1.88	11.2	2.12
Topic Development	7.0	2.01	8.3	1.96	7.7	1.99
English Standards	5.8	1.30	6.1	1.29	6.0	1.22
Total Score	12.8	3.09	14.4	3.12	13.7	2.99
Passage Length	305	141.1	445	258.7	332	159.2
WPM	22.1	9.04	24.2	8.48	24.8	11.76

Figure 2: Mode of administration effect on MCAS Language Arts scores - grade 4.

To examine whether these differences were statistically significant, a one-way analysis of variance was performed. Table 13 indicates that the group means for Topic Development, Total Score and Passage Length did differ significantly among the three groups. Scheffe post-hoc comparisons were then performed for these three variables. Table 14 indicates that none of the differences between paper and AlphaSmarts were statistically significant. However, the differences between paper and computer were all statistically significant. Specifically, the mean scores for Topic Development and Total Score were significantly higher for students who performed the composition item on computer as compared to performance on paper. In addition, students who worked on computer wrote longer passages than students who wrote on paper or on AlphaSmarts.

Table 13:
Mode of Administration ANOVA

	Mean Square			
	Between	Within	F	Sig.
Mid-term Grade	3.36	3.67	.92	.40
Topic Development	20.94	3.94	5.31	.006
English Standards	1.18	1.16	.73	.48
Total Score	31.91	9.39	3.40	.03
Passage Length	274,892	37,359	7.36	.001
WPM	99.19	98.26	1.01	.36

Table 14:
Scheffe Multiple Comparisons for Topic Development, Total Score and Passage Length

	Paper vs Computer			Paper vs. AlphaSmart			Computer vs. AlphaSmart		
	Mean Diff	Std. Error	Sig.	Mean Diff	Std. Error	Sig.	Mean Diff	Std. Error	Sig.
Topic Development	1.30	.40	.006	.68	.39	.28	.62	.39	.29
Total Score	1.6	.62	.036	.88	.61	.35	.73	.60	.49
Passage Length	140	38.8	.002	27.5	38.3	.77	112	38.1	.014

Although the computer group scored significantly higher than the paper group, the computer group also had slightly higher mid-term grades. To control for differences in mid-term grades, Topic Development, English Standards and Total Score were each regressed on mid-term grades and on group membership. For these regressions, two dummy variables were created for group membership. The first, called Computer, was coded as 1 for students who performed the test on computer and 0 for all other students. The second dummy variable, called Alpha, was coded 1 for students who performed the test on an AlphaSmart and 0 for all other students. Table 15 indicates that after controlling for differences in mid-term grades, performing the composition item on computer still had a significant effect on students' Topic Development and Total scores. It is interesting to note that although students working on computer had access to a spell-checker, this access did not result in significantly higher English Standards scores. Access

to a computer, however, did enable students to write longer passages and, in turn, receive significantly higher scores for Topic Development. Conversely, access to an AlphaSmart did not have a significant effect on student scores.

Table 15:
Composition Scores Regression Analyses for Grade 4

Topic Development					
R=.54	**B**	**Std. Error**	**Beta**	**T**	**Sig.**
Constant	1.53	.83			
Mid-Year Grade	0.51	.07	.48	6.88	<.001
Computer	1.17	.35	.37	3.34	.001
Alpha	0.42	.35	.10	1.21	.23
English Standards					
R=.54	**B**	**Std. Error**	**Beta**	**T**	**Sig.**
Constant	2.06	.52			
Mid-Year Grade	0.35	.05	.53	7.67	<.001
Computer	0.21	.22	.08	0.96	.34
Alpha	0.02	.22	.01	.09	.93
Total Score					
R=.57	**B**	**Std. Error**	**Beta**	**T**	**Sig.**
Constant	3.58	1.24			
Mid-Year Grade	0.86	.11	.53	7.81	<.001
Computer	11.37	.52	.21	2.64	.009
Alpha	0.44	.52	.07	0.85	.40

Examining Keyboarding Speed and Mode of Administration Effect

The regression analyses presented above (Tables 11 and 15) indicate that mode of administration had a significant effect on students' performance in grade eight and that administration on computer had a significant effect on students performance in grade four. To test whether the effect of mode of administration varied for students with different levels of computer skill, students' WPM was used to form three groups. The first group contained students whose WPM was .5 standard deviations below the mean. The second group contained

students whose WPM was between .5 standard deviations below the mean and .5 standard deviations above the mean. The third group contained students whose WPM was .5 standard deviations above the mean. For each group, the composition total score was regressed on mid-term grades and group membership. Below, these three sub-group regressions are presented separately for grade eight and grade four.

Grade 8. In grade eight, "slow" keyboarders were defined as students whose keyboard speed was lower than 23.8 words per minute. "Average" keyboarders typed between 23.8 and 32.4 wpm. And "fast" keyboarders typed more than 32.4 wpm. Table 16 displays the results of the three separate regressions. For all sub-groups, performing the composition item on computer had a positive effect on students' scores. However, the size of the effect increased as keyboarding speed increased. For slow keyboarders, the effect represented an increase of about .33 standard score points. For fast keyboarders, the effect was about 1.7 times larger.

Table 16:
WPM Sub-group Regression Analyses for the Paper versus eMate Experiment

WPM<23.8 N=22					
R=.48	B	Std. Error	Beta	T	Sig.
Constant	-1.71	6.86			
Mid-Year Grade	0.17	0.08	.43	2.09	.049
Group	1.38	0.84	.33	1.63	.119
23.8>WPM<32.4 N=33					
R=.65	B	Std. Error	Beta	T	Sig.
Constant	-8.93	5.98			
Mid-Year Grade	0.26	0.07	.52	3.82	.001
Group	2.07	0.78	.37	2.67	.012
WPM>32.4 N=26					
R=.71	B	Std. Error	Beta	T	Sig.
Constant	-5.69	6.61			
Mid-Year Grade	0.22	0.07	.44	3.02	.006
Group	2.68	0.71	.55	3.77	.001

Grade 4. In grade four, "slow" keyboarders were defined as students whose keyboard speed was lower than 18.7 words per minute. "Average" keyboarders typed between 18.7 and 28.7 wpm. And "fast" keyboarders typed more than 28.7 wpm. Table 17 displays the results of the three separate regressions. For fast keyboarders, performing the composition item on computer or on an AlphaSmart

had a moderate, positive effect on students' scores. The size of the effect was substantially smaller for average keyboarders who composed on a computer and was slightly negative for students who composed on an AlphaSmart. And for slow keyboarders, using a computer had virtually no effect while using an AlphaSmart had a negative effect. Note that the effects were only statistically significant for fast keyboarders.

Table 17:
WPM Sub-group Regression Analyses for the Paper versus Computer Experiment

WPM<18.7 N=52					
R=.66	B	Std. Error	Beta	T	Sig.
Constant	1.69	1.76			
Mid-Year Grade	1.00	0.17	.65	5.98	<.001
Computer	0.18	0.77	.03	0.24	.82
Alpha	-.65	0.69	-.11	.94	.35
18.7<WPM<28.7 N=56					
R=.30	B	Std. Error	Beta	T	Sig.
Constant	10.0	2.06			
Mid-Year Grade	0.36	0.19	.26	1.97	.05
Computer	0.63	0.90	.12	.70	.49
Alpha	-.32	0.93	-.06	.34	.73
WPM>28.7 N=44					
R=.64	B	Std. Error	Beta	T	Sig.
Constant	3.11	2.72			
Mid-Year Grade	0.89	0.23	.49	3.94	<.001
Computer	2.25	0.89	.36	2.54	.02
Alpha	2.22	0.89	.36	2.48	.02

Discussion

The two experiments described here extend the work of Russell (1999) and Russell and Haney (1997) in two important ways. In addition to examining the mode of administration effect in grade eight, these studies examine the effect in grade four. In many state testing programs as well as the National Assessment of Educational Progress and international studies such as the Third International Mathematics and Science Study, these two grade levels are commonly tested. Thus, it is important to understand the extent to which mode of administration affects the performance of students in these commonly tested grade levels. Second, these two experiments introduce a third mode of administration that provides schools and testing programs with a cheaper word processing option, namely AlphaSmarts and eMates.

As in the two previous studies, the fourth grade study presented above found that students who wrote their compositions on computer produced longer responses that received higher scores. This computer effect was statistically and practically significant. In addition, substituting an eMate for a desktop computer also had a positive effect on students' performance in grade eight. Substituting an AlphaSmart for a computer in grade four, however, had a less pronounced effect.

Across both studies, the relationship between keyboarding speed and the mode of administration effect was inconsistent. As Russell (1999) found, the fourth grade study presented here indicates that students need to have sufficient keyboarding speed before the mode of administration effect becomes meaningful. In grade eight, however, this pattern did not emerge. Although the effect was largest for fast keyboarders who produced their composition on an eMate, the effect was positive and of practical significance at all levels of keyboarding speed.

One reason why this finding differs from that reported by Russell (1999) may stem from the relatively high level of keyboarding speed for most students included in the grade eight experiment. Whereas Russell (1999) reported an average keyboarding speed of about 15 words per minute, average speed for the grade eight students included in this study was nearly twice as fast. Moreover, the average speed of students included in the "slow" keyboarding group for the eighth grade study exceeded the cut point for the high keyboarding level in Russell's (1999) study. It appears that once eighth grade students achieve keyboarding speed of roughly 20 to 24 wpm, the mode of administration effect becomes meaningful. Fourth grade students, however, may need better keyboarding skills before the effect occurs. In addition, when students use portable writing devices rather than desktop computers, the size of the effect may increase.

Limitations

These two experiments focused on students in grades four and eight attending school within a single district. This district tends to perform well above the state average on standardized and state level tests. In the studies presented here, very few students performed at low levels on the composition item. As a result, it was not possible to examine the mode of administration effect across the full range of performance levels.

Similarly, students included in this study were generally accustomed to working with computers. The relatively high level of keyboarding speed complicated efforts to examine the mode of administration effect at low levels of keyboarding

speed in grade eight. Additionally, students' familiarity with computers prevented an examination of the mode of administration effect for students who are not accustomed to working with computers. Despite this high level of access to computers within Wellesley's schools, it should be noted that approximately 25% of the school districts across the state have a better student to computer ratio than does Wellesley.

Finally, due to a limited number of computers in grade four, students who composed their first draft on an AlphaSmart were required to wait one school day before they could edit their final draft on computer. Students who worked on paper or who composed their first and second drafts entirely on computers completed their compositions during two consecutive one-hour blocks. It is possible that the extended time between composing the first and second drafts may have affected the performance of fourth grade students who worked on AlphaSmarts.

Implications

This series of studies provides further evidence that students accustomed to writing on computers or on eMates perform significantly better when open-ended tests are administered on a computer. For the MCAS Language Arts test, this improved performance translates into approximately two points on the composition item. But in addition to the composition item, the MCAS Language Arts test contains four short answer open-ended items each worth four points. Assuming that the effect found for the composition item and that the effect reported on the shorter open-ended items by Russell (1999) holds across all Language Arts open-ended items, students accustomed to writing with computers may perform better by an additional two points on these open-ended items. Across all short and extended open-ended MCAS items, the mode of administration effect may result in an increase of four raw score points if students were allowed to compose responses on a computer. An increase of four raw score points translates into between four and eight point scale score points, depending upon where on the scale a student's score resides. This score increase may be even larger for fast keyboarders in grade four. Clearly, as state testing programs such as MCAS begin or continue to use test scores to make critical decisions about graduation and promotion, steps should be taken that allow students who are accustomed to working on computers to perform at their highest level.

School level effect. Within Wellesley, eliminating the mode of administration effect for both the composition item and the four shorter open-ended items would have a dramatic impact on district level results. As Figure 3 indicates, based on last years

(1999) MCAS results, 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, Figure 4 shows that in grade eight, four percent of students would move from the "Needs Improvement" category to the "Proficient" category and that 13% more students would be deemed "Advanced." As Figure 5 displays, within one elementary school (Bates), the percentage of students performing at or above the "Proficient" level would nearly double from 39% to 67%.
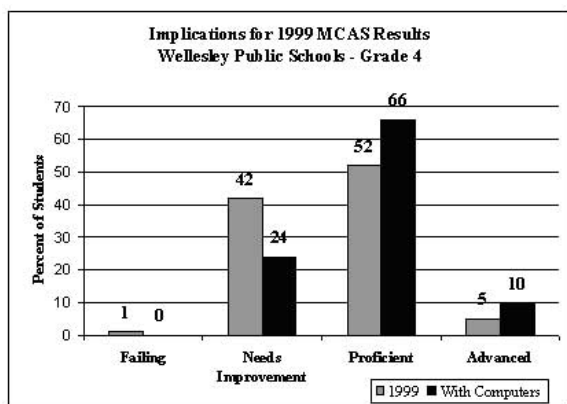
Figure 3. Mode of administration effect on grade 4 1999 MCAS results.
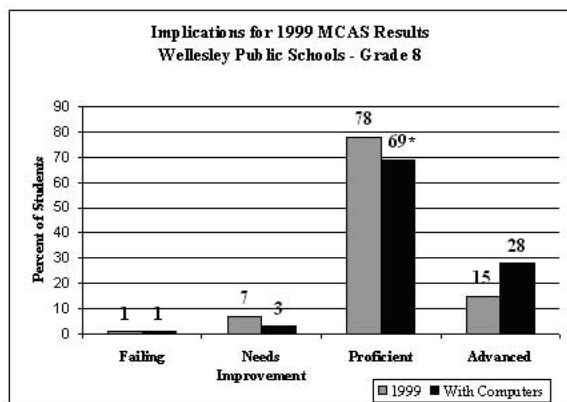
Figure 4. Mode of administration effect on grade 8 1999 MCAS results.

The mode of administration effects reported here and in previous studies (Russell, 1999; Russell and Haney, 1997) highlight a complicated challenge testing programs must overcome as they continue to use tests containing open-ended items to make inferences about student and school achievement. To reduce the mode of administration effect, state testing programs should consider allowing students to select the mode by which open-ended responses are composed. For the past decade, the Province of

Alberta has employed this strategy for its graduation testing program (Alberta Learning, 2000). Over the past five years, the province has seen the percentage of students opting to perform the English, Social Studies, Biology and French tests on computer increase from 6.7% in 1996 to 24.5% in 2000. Within high schools, the percentage of students opting to perform the test on a computer ranges from 0 to 80% (Sakyi, 2000).
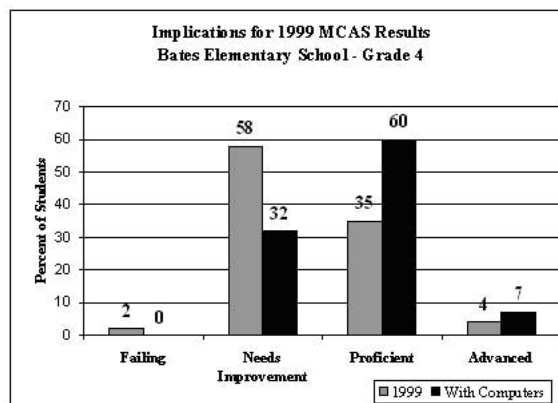

Figure 5. Mode of administration effect on Bates Elementary School 1999 MCAS results.

Although this approach adds to the complexity to test administration procedures (see Russell & Haney, 2000 for a more detailed review), providing students the option of working on paper, on computer, or on a portable writing device would create writing conditions that more closely reflect normal practice. In turn, the authenticity of testing would increase, students would be able better able to demonstrate their best work, and more valid measures of student writing would be available to judge student and school achievement.

References

Alberta Learning. (2000). Directions for Administration, Administrators Manual, Diploma Examination Program.

Arnold, V., Joyner, R. L., Schmidt, B. J., & White, C. D. (1997). Establishing electronic keyboarding speed and accuracy timed-writing standards for postsecondary students. *Business Education Forum, 51*(3), p. 33-38.

Bunderson, C. V., Inouye, D. K. & Olsen, J. B. (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.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd ed.) Hillsadale, NJ: Erlbaum.

Council of Chief State School Officers (1998). *Key State Education Policies on K-12 Education: Standards, Graduation, Assessment, Teacher Licensure, Time and Attendance.* Washington, DC.

Glass, G. & Hopkins, K. (1984). *Statistical Methods in Education and Psychology.* Boston, MA: Allyn and Bacon.

Kleiman. 2000. Myths and Realities about Technology in K-12 Schools. In D. Gordon (ed.), *The Digital Classroom.* Cambridge, MA: Harvard Education Letter.

Massachusetts Department of Education. (1999a). Release of Spring 1999 Test Items. Malden, MA. (http://www.doe.mass.edu/mcas/99release.html).

Massachusetts Department of Education. (1999b). 1998 MCAS Technical Report Summary. Malden, MA.

Mead, A. D. & Drasgow, (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114* (3), 449-58.

Meyer, L., Orlofsky, G., Skinner, R., & Spicer, S. (2002). The state of the states. *Education Week, 21*(17), 68-169.

NTS Computer Systems. (1998). *1998 Annual Report.* British Colombia, CN: NTS Computer Systems, LTD.

Powers, D., Fowles, M, Farnum, M, & Ramsey, P. (1994). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement, 31*(3), 220-233.

Robison, J. W., Erickson, L. W., Crawford, T.J. & Ownby, A. C. (1979). *Typewriting: Learning and Instruction.* Cincinnati: South-Western Publishing Company.

Russell, M. (1999). Testing Writing on Computers: A Follow-up Study Comparing Performance on Computer and on Paper. *Education Policy Analysis Archives, 7*(20). Available: http://epaa.asu.edu/epaa/v7n20/.

Russell, M. & Haney, W. (1997). Testing writing on computers: an experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives, 5*(3). Available: http://epaa.asu.edu/epaa/v5n3.html.

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

Russell, M. & Plati, T. (2001). Mode of Administration Effects on MCAS Composition Performance for Grades Eight and Ten. *Teachers College Record*, [Online]. Available: http://www.tcrecord.org/Content.asp?ContentID=10709.

Russon, A. R. and Wanous, S. J. (1973). *Philosophy and Psychology of Teaching Typewriting.* (2nd ed.) Cincinnati: South-Western Publishing Company.

Sakyi, A. (2000). Word Processing in Humanities Diploma Examinations. Personal correspondence from the Alberta Learning, Student Evaluation Branch.

West, L. J. (1968) The vocabulary of instructional materials for typing and stenographic training - research findings and implications. *Delta Pi Epsilon Journal, 10*(3), 13-125.

West, L. J. (1983). *Acquisition of Typewriting Skills,* (2nd ed.), Indianapolis: The Bobbs-Merrill Company, Inc.

Wolf, F. (1986). *Meta-analysis: Quantitative methods for research synthesis.* Sage University series on quatitative applications in the social sciences, series no. 07-059. Newbury Park, CA: Sage.

Does it Matter with What I Write? Comparing Performance on Paper, Computer and Portable Writing Devices

2002 Article Citation
Russell, M., & Plati, T. (2002, May 29). Does it matter with what 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/

Author Notes
Michael Russell
Technology and Assessment Study Collaborative Center for the Study of Testing, Evaluation and Educational Policy
Boston College
russelmh@bc.edu

Tom Plati
Director of Libraries and Educational Technologies
Wellesley Public Schools
tom_plati@wellesley.mec.edu

2015 Article Citation
Russell, M., & Plati, T. (2002). Does it matter with what I write? Comparing performance on paper, computer and portable writing devices. *Current Issues in Education*, *5*(4). Retrieved from http://cie.asu.edu/ojs/index.php/cieatasu/article/view/1621

Current Issues in Education
Mary Lou Fulton College of Education
Arizona State University