

The Effect of a Word Processor as an Accommodation for Students with Learning Disabilities

Cassie L. Berger & Larry Lewandowski

Syracuse University | United States

Abstract: The effects of writing format (handwritten (HW) versus word processor (WP)) were examined in a sample of college students with and without learning disabilities (LD). All students wrote two essays, one in each format, scored for quality and length. Groups did not differ in age, gender, ethnicity, mathematical calculation, writing fluency, essay length or essay quality. The “interaction hypothesis” was not supported, in that the use of a word processor as a writing accommodation did not provide a differential boost to students with LD. Both groups produced longer essays in the WP versus HW condition. The best predictor of essay quality was essay length regardless of writing format. Most students in each group preferred the WP format. Interestingly, a smaller percentage of students in the LD group (72%) than NLD group (91%) used the available time for writing.

Keywords: accommodations, learning disabilities, college students, writing, word processor



Berger, C. L., & Lewandowski, L. (2013). The Effect of a Word Processor as an Accommodation for Students with Learning Disabilities. *Journal of Writing Research*, 4(3), 261-280.
<http://dx.doi.org/10.17239/jowr-2013.04.03.2>

Contact: Cassie Berger, Syracuse University/Dept. of Psychology, 430 University Avenue, Syracuse, NY 13244 | United States – cassieberger@gmail.com.

Copyright: Earli | This article is published under Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported license.

It is well recognized that students with LD form the largest disability group receiving services in high schools and colleges (U. S. Department of Education, National Center for Education Statistics, 2011). Within the group of students with LD, about 80-90% has language-based disabilities, especially in reading and writing (Hughes & Osgood Smith, 1990). It follows that many of these students have difficulties on tests that require challenging reading and writing components, prompting these students to seek test accommodations for classroom and high stakes exams. Some of the most frequently allowed test accommodations include: dictated response (scribe), large print, Braille, reader, and extended time (Bolt & Thurlow, 2004). Another common test accommodation for LD has been the use of a computer for writing. However, there has been very limited research studying the effectiveness of a computer as a test accommodation for written expression.

1. Introduction

The addition of essay sections on high stakes exams has been a relatively recent trend. The Scholastic Abilities Test (SAT), Graduate Record Exam (GRE), Law School Admission Test (LSAT), and Test of English as a Foreign language (TOEFL), to name a few, each require examinees to write essays under specified time constraints. As successive generations of students perform less handwriting and more word processing, standardized exams and classroom exams incorporating computer-assisted writing formats are more prevalent. Some high stake exams, like the GRE, have become computer based; others, such as some Law Bar exams, allow the use of laptop computers during exams (Ewing et al., 2005). Yet, the SAT as well as most school districts and universities have computer/WP listed on their menus of test accommodations available to students with disabilities. Until WPs become the standard for tests of writing, students with LD will continue to apply for the use of a WP as a test accommodation (Li & Hamel, 2003). In England, for example, the government routinely purchases laptop computers for university students with LD, and many of their written exams are taken on the computer regardless of their particular diagnosis (i.e., dyslexia).

1.1 Accommodations

Test accommodations for students with LD have become somewhat controversial. Researchers, educators, and even politicians have questioned the validity and fairness of test accommodations for students with LD (Gordon, Lewandowski, & Keiser, 1999; Lovett, 2010; Lovett & Lewandowski, 2006; Ranseen & Parks, 2005). Currently, there is a lack of research-based guidance for the use of certain test accommodations in certain situations for certain persons. Whereas there may be an obvious need for a blind person to take a test in Braille, it is not as straightforward to conclude that a student with diagnosed LD should receive double time on a test relative to his or her peers. Rather than make such decisions based on a philosophical position, it would seem that the

field would benefit from empirical answers to the question of validity for specific test accommodations.

Phillips (1994) argues that an appropriate test accommodation should mitigate performance obstacles of students with disabilities (e.g., large print for a student with visual limitations), while having little to no effect on the performance of typical students. Fuchs and Fuchs (2000b) have referred to this as a “differential boost,” since the scores of students with disabilities receive a greater boost from the accommodation. Sireci, Scarpati, and Li (2005) pointed out that a differential boost can occur even if nondisabled students benefit substantially from the accommodation, so long as students with disabilities receive an even greater benefit—that is, so long as there is a statistical interaction between disability status and the presence or absence of the accommodation (*interaction hypothesis*). However, other scholars (e.g., Zuriff, 2000) have argued that to be a fair accommodation, nondisabled students—who typically do not have access to the accommodation—should not benefit from it. According to Zuriff, the predicted interaction is that the performance of the students with no learning disabilities (NLD) will not change as accommodations are implemented, whereas the performance of students with LD will improve. In other words, the groups will differ under standard conditions, but be relatively equivalent under accommodated conditions.

1.2 Word Processors as Test Accommodations

As discussed earlier, there are many college students with LD and they account for a significant portion of test accommodation requests (Henderson, 2001). The most common accommodations for college students include the use of extended time, special test setting, computers, calculators, readers, and large print (Sireci et al., 2005). Research has shown that the cognitive demands of sentence composition and writing impacts performance (Van Waes, Leijten, & Quinlan, 2010), therefore, requests for accommodations, especially in regard to writing, have increased. Among college students with LD experiencing writing difficulties, word processors, speech recognition systems, speech synthesis systems, and multimedia technology are the most frequently used tools (Li & Hamel, 2003). In the last decade, use of computers in schools has risen along with the use of computers as tools for assessment. It has been argued that, “technology-based intervention and assessment refers to using the computer or other expert systems as the medium to provide instruction and monitor students learning” (Maccini, Gagnon, & Hughes, 2002, p.248). According to the National Center for Education Statistics (2011), the majority of all universities (public, private, two-year, and four-year) permit the use of assistive technology (including WP) as testing accommodations for college students with LD. As computers in elementary, middle, high school, and college environments are becoming more widely used, accepted, and required, it is increasingly important that we evaluate the relative benefits of computerized writing as either a test accommodation for some, or an applied technology for everyone.

1.3 Studies Comparing Handwritten and Word Processed Essays

Word processors serve as tools for writing and learning as well as accommodations for students with LD with writing difficulties. A WP can accommodate students with difficulties in spelling, grammar, organization and penmanship. To our knowledge there are no studies that examine the differential boost/interaction hypothesis with regard to the WP test accommodation. In other words, there is no evidence showing that use of a WP boosts performance on writing assignments for students with LD significantly more than for NLD students.

With regard to nondisabled students, Russell and Haney (1997) examined the two writing formats (computer and paper-and-pencil) in a sample of students in grades 6-8. These researchers found that written outputs such as essay length and paragraph organization were improved with the use of a WP. In addition, researchers in Hong Kong comparing HW versus WP compositions of secondary school students found that overall, students who used a computer composed better essays (Lam & Pennington, 1995). Similarly, meta-analytic studies have indicated that use of a WP increased length and improved quality of writing (Bangert-Drowns, 1993), and WP can be an effective intervention for increasing writing competence (Graham & Perin, 2007). More specifically, Morphy and Graham, 2012 conducted a meta-analysis based on 27 studies with weaker writers in grades 1-12. These researchers found moderate to large effects for the use of a word processor versus handwriting in the following areas: writing quality ($d = 0.52$), length ($d = 0.48$), development/organization of text ($d = 0.66$), mechanical correctness ($d = 0.61$), motivation to write ($d = 1.42$), and preferring word processing over writing by hand ($d = 0.64$). In addition Graham, Harris, and Hebert (2011) found that when students were experienced in the use of a word processor, a statistically significant effect for quality was found ($d = 0.54$) supporting word processing. On the other hand, when investigating differences in scores for paper and computer versions of a writing test for 8th grade students, Horkay, Elliot, Bennett, Allen, Kaplan, and Yan (2006) found that there were no significant differences in the mean essay scores on the paper and computer based assessments. Similarly, a group of college students at the University of Edinburgh were asked to complete mock examinations in either HW or WP formats and no significant differences were found in regard to essay length or quality score (Mogey, Paterson, Burk, & Purcell, 2010). Van Waes and Schellens (2003) studied the effect of writing mode on experienced adult writers. They found differences between profiles of writers based on HW versus WP mode as well as other constraints of the writing environment. Additionally, studies comparing HW and WP essay formats found evidence of rater bias. Studies show a positive bias towards HW essays, finding that these essays received higher ratings than WP essays (Arnold, Legas, Obler, Pacheco, Russell, and Umbdenstock, 1990; Bridgeman and Cooper, 1998; & MacCann, Eastment, and Pickering, 2002).

Few studies have compared HW and WP testing formats within a sample of students with LD and only three studies have sampled students with and without LD evaluating HW and WP essays. MacArthur and Graham (1987) explored how different

conditions of text production influence the writing process and products of a heterogeneous group of students with LD. Each participant composed a story in HW, WP and dictation (D) from three different picture prompts. Interestingly, results show that there were no significant differences between HW and WP formats. Mechanical errors and number of revisions were similar for both formats. According to the rate measure, the HW format appeared to be twice as fast as the WP format. This study suggests that WP by itself may not have a major impact on the writing of students with LD, and actually may demand more attention than HW to the mechanical aspects of text production. Similarly, in a study comparing paper and pencil and WP use with elementary students with LD, MacArthur (1996) again found no significant differences between the formats.

Hollenbeck, Tindal, Stieber, and Harris (1999) compared middle school students with disabilities in special education classes to students in regular education classes on HW and WP essays. All essays were originally HW and later transcribed into WP format and scored by two raters on six traits that included: ideas and content, organization, voice, sentence fluency, word choice, and conventions. Results were consistent with previous studies of NLD groups showing that when original HW essays were typed and scored, the original HW essays were rated significantly higher on four out of the six traits that included content, organization, ideas, and conventions. However, it is important to note that these researchers did not test for differences between the special education and regular education students in terms of essay quality.

Berninger, Abbott, Augsburger, and Garcia (2009) compared elementary students with LD in spelling and handwriting to NLD students. Students were asked to compose letters, sentences, and essays in both HW and WP formats. These researchers found that both groups took longer to compose sentences and essays in the WP format. Additionally, both LD and NLD elementary school students wrote longer essays and more complete sentences in the HW than the WP condition.

Recently, Gregg, Coleman, Davis, and Chalk (2007) compared college students with ($n = 65$) and without dyslexia ($n = 65$) on essay writing. Researchers investigated whether there is a difference in the quality score of essays in handwritten, typed, and typed/edited formats as well as the influences of spelling, handwriting, fluency, and vocabulary complexity on the quality scores of these students. All of the participants received 30 minutes to complete an expository essay. Essays were scored in original handwritten format, typed format, and a format that was typed and edited for spelling mistakes, agreement problems, and punctuation errors. There were no significant differences between the different formats for the participants with and without dyslexia. Quality scores were significantly lower for students with dyslexia compared to students without dyslexia on all three formats. Additionally, there was a high correlation between verbosity (quantity of writing) and quality. For both groups of participants, quality scores were not significantly different between the handwritten and typed essays. Interestingly, out of the students with dyslexia, only 71% completed their essays

compared to the students without dyslexia where 91% completed their essays in the allotted 30 minutes.

1.4 Summary

There is some evidence that students with LD (mixed types) perform lower than NLD students on various writing tasks, yet no conclusive results showing that students with LD perform better with accommodations on such tasks. There is evidence that HW essays are rated higher than WP essays, yet other evidence showing that WP essays produce more writing of higher quality. Questions remain as to the benefits of a WP, as well as its validity as an accommodation for students with LD. The following study examined the interaction hypothesis in a design that compared LD and NLD college students on essay quality and length for both WP and HW writing formats. The hypothesis predicted that LD students would perform less effectively than NLD students in the HW format, and perform comparably to peers in the WP (accommodated) format.

2. Method

2.1 Participants

Table 1. Group Intercorrelations of Writing Variables (n=98)

Variable	NLD	LD	Total
	(n=68)	(n=30)	(n=98)
Male	36(53.3%)	16(52.2%)	52(52.5%)
Female	32(46.7%)	14(47.8%)	46(47.5%)
Age	19.1	19.3	19.2
Caucasian	36 (53.3%)	26 (86.7%)	62 (63.6%)
Afr-Amer.	5 (7.2%)	0 (0.0%)	5 (5.1%)
Asian-Amer.	18 (26.1%)	0 (0.0%)	18 (16.9%)
Latin-Amer.	4 (5.8%)	2 (6.7%)	7 (7.1%)

Participants included 98 undergraduate college students ages 17-24 ($M = 19.04$), 47% female, from a large, private university in central New York (see Table 1). The majority of participants were enrolled in an introductory psychology class and were offered one hour of course credit for participation in the experiment. Additional participants were recruited from other undergraduate psychology courses offered on campus and paid \$20 for their time. The majority of the participants were freshman (63%), then sophomores (22%), juniors (11%) and seniors (4%). The ethnicity of the sample was predominately Caucasian (64%), followed by Asian (18%), Latino (6%), African-American (5%), and Other (7%).

Sixty-eight participants were identified as NLD (36 males, 32 females) and 30 participants were identified with a professional diagnosis of LD according to the Office of Disability Services (ODS), (16 males, 14 females). Selection for the LD group required documentation from the ODS that indicated the student was receiving special services due to a learning disability. The ODS review policy is based on documentation including a current evaluation conducted by a qualified evaluator (within 3 years at the time of this study). This comprehensive assessment must include a diagnostic interview, aptitude/IQ testing, academic achievement testing, information processing testing, and a specific DSM IV-TR diagnosis. Students with LD were approved by ODS to receive test accommodations if they had a recent psychological assessment that met DSM IV-TR criteria. Students with medical or other psychological diagnoses that may influence writing or test behavior were excluded. Students in the LD group had documented learning disabilities in at least one of three areas including reading, writing, or processing speed. It was inferred that students with such disabilities could have difficulties on timed tests of essay writing. About 40% of students with LD provided documentation in at least two of the three above-mentioned disability areas. Ninety-two percent of participants in the LD group were receiving a WP as a test accommodation from the university, indicating that regardless of their area of impairment they were allowed to write exams on a computer.

Unfortunately, in the current study we were unable to review psycho-educational evaluation profiles for these students and we could not assess their extent of need for a word processor. We simply accepted the common practice, that as students with an acknowledged disability, they were granted the use of this accommodation for exams that involved writing. It appears that most schools and many test agencies are rather liberal in allowing requested accommodations once a student's disability status has been verified.

As one might expect, the two groups differed in grade point average, with LD ($M = 3.0$) reporting lower GPA than NLD ($M = 3.3$) students, $t(1,96) = 2.5, p < .05$. The two groups were not significantly different with regard to sex, $\chi^2(1, N = 98) = .01, p = 0.55$, ethnicity (percentage of Caucasian students vs. non Caucasian students), $\chi^2(1, N = 98) = .67, p = 0.17$, year in school, $\chi^2(3, N = 98) = 7.71, p = 0.05$, or age, $t(1, 96) = -1.47, p = .15$.

2.2 Measures and Materials

Questionnaire

A questionnaire was developed in order to: 1) provide descriptive demographic information; 2) document a history of accommodations, writing difficulties, tutoring in writing, and time spent using a computer; and 3) elicit preference ratings for WP and HW formats.

Computer

A Dell Latitude C840 laptop computer equipped with a Word Pad software program was used in order to compose the WP essays. Word Pad does not automatically correct for spelling or grammar, allowing for both WP and HW essays to be recorded in their original state.

Number-Writing Task (NWT)

The NWT was developed in order to measure speed of typing. The task was completed on a computer by typing the word forms of numbers such as “one, two, three...twenty-one” for one minute. Each character typed was counted separately and totaled to measure speed of typing. Errors, spaces, and commas were not counted. The measure was designed to examine possible differences in typing speed between the LD and NLD groups and to determine the relationship between typing speed and writing length. The correlation between the NWT and length for the combined groups (LD and NLD) was .57.

Woodcock Johnson III Tests of Achievement Writing Fluency Subtest (WJ-III)

The Writing Fluency Subtest measures the ability to quickly formulate and write simple sentences about pictures using three given words (e.g., good, cake, is). It was used to assess potential differences between groups in fluent production of writing (assuming that the LD group would perform lower than the NLD group). Significant differences on the Writing Fluency subtest were found in a sample of university students with and without LD, providing support for the use of this measure in the classification of students with LD, $t(159) = -11.4, p < .001$. Test-retest reliability for the Writing Fluency Subtest is reported as $r = .88$ (Woodcock, McGrew, & Mather, 2001).

Woodcock Johnson III Tests of Achievement Calculations Subtest (WJ-III).

The Calculations Subtest of the WJ-III measures performance in mathematical calculations. The Calculations Subtest was used to measure skills that were assumed to be similar across LD and NLD groups (MacArthur & Cavalier, 2004). Equivalence on this subtest would suggest that differences in writing skills were not attributed to overall low achievement. Reliability on this subtest is reported as ($r = .85$; Woodcock, McGrew, & Mather, 2001).

Wechsler Individual Achievement Test – Second Edition, Written Expression Subtest (WIAT-II).

The WIAT-II is a comprehensive, individually administered test for assessing the achievement of children, adolescents, college students, and adults. Specific subtests can be administered in groups as well. The WIAT-II provides normative data for college students in grades 13 through 16 and adults age 17 through 85 years. Importantly, there is a supplemental guide for administering and scoring the essay for college students and adults. For the purpose of this study, the essay portion of the Written Expression Subtest was administered. Written Expression assesses the writing process, specifically, the

mechanics, organization, theme development, and vocabulary of essay writing. There are guidelines for assessing word count and a detailed rubric to calculate a quality score. Two prompts, Prompt A and B, were used as stimuli for essay writing. Each prompt asked students to take a position on a particular issue and elaborate on three supporting arguments. Prompt A asked about free college tuition whereas Prompt B asked about daylight savings time. This format is similar to essay prompts on high stakes exams such as the SAT and GRE. Scoring was assessed for quality and word count (quantity/length) using guidelines from the Written Expression Subtest of the WIAT-II. Overall quality of writing was evaluated according to dimensions outlined in the test manual that include point of view, organization of ideas, theme development, vocabulary, and content. There was one quality score for each essay written. In the college sample, quality scores show high test-retest and interscorer agreement ($r = .77$ and $.81$; Wechsler, 2002). Total word count or length was measured at completion of the essay. Rate was determined by dividing length at five minutes by five (creating a words per minute rate of production score). Standard time on the WIAT-II essay is 15 minutes. The time was decreased from 15 to 10 minutes in order to present a more time-sensitive task similar to high-stakes exams. Completion time for both the HW and WP conditions was recorded.

2.3 Procedure

The study was conducted in a small, quiet room that accommodated up to ten students. Most sessions were conducted in small groups, though four students were tested individually. Participants signed an informed consent form then completed the NWT to measure typing speed. Following the NWT, participants were asked to write two essays using two different stimulus prompts (A & B) from the Written Expression Subtest of the WIAT-II. Materials and administration were the same for all participants. The sessions were counterbalanced for order of response format (WP vs. HW) and stimulus prompt (A vs. B). For the WP conditions, students typed on the laptop computer and for the HW conditions students wrote on lined paper with a pencil. To increase motivation and investment in the task, participants were told that they should do their very best as they were receiving "a little more time than usual" to complete their essays. Participants were also informed that essays were scored based on ideas, clarity, and arguments rather than on mechanics.

At the 5-minute mark, students were given notification that there were five minutes left, and asked to underline the word that they were writing or typing at that time. At 10-minutes, students were asked to stop writing. If participants finished before the 10-minute time allotment, they were asked to raise their hand to let the examiner record the time. Essays composed on the WP were saved by code number on a USB drive, and later transferred to a computer database.

Following the Written Expression Subtest of the WIAT-II, students were asked to complete the Writing Fluency and Calculations Subtests from the WJ-III in a fixed order.

Finally, students were administered the questionnaire. Total testing time was approximately 50 minutes.

Prior to scoring, all HW essays were typed so raters were not able to determine whether the essay had originally been handwritten or typed by the participants. The HW essays were typed verbatim without correcting for spelling or grammar. All essays were presented to raters typed, double-spaced, in Times New Roman and 12-point font. The dependent variables were quality, length of essay, and rate at both five minutes and completion (maximum time 10 minutes).

2.4 Interscorer Agreement

A primary researcher scored all of the essays for length and quality. Two undergraduate researchers were trained to 100% agreement on length and 80% agreement on quality with the primary researcher. Thirty percent of the WIAT-II, Written Expression essays were randomly selected and re-scored by two out of three trained raters for total words written and quality. The inter-rater reliability coefficient for number of words written was 1.0 (100% agreement). Raters also assessed each essay with a quality score that ranged from zero to six according to the WIAT-II Written Expression scoring guidelines manual. The two quality scores for each double-scored essay were then analyzed for correlations between scores. The inter-rater reliability coefficient for the quality score was $r = .80$.

2.5 Relationships of Dependent Measures

Correlations were computed among the various performance measures for both LD and NLD groups separately and combined (see Tables 2 and 3). Interestingly, correlations on the writing variables were similar for both groups. NLD and LD groups showed strong correlations between WP quality score and WP length (NLD: $r = .67$, LD: $r = .76$) and HW quality score and HW length (NLD: $r = .54$, LD: $r = .59$). This suggests that longer essays, for both groups, in either format, tend to be associated with higher quality scores. In the NLD group, performance measures (writing fluency and number writing) correlated significantly (all r -values $> .40$) with dependent measures (length and quality scores) in the WP condition. Writing fluency also correlated significantly with length ($r = .36$) and quality ($r = .27$) scores in the HW condition for the NLD group. In the LD group, the correlations were slightly less between writing fluency and length in the WP ($r = .28$) and HW ($r = .26$) conditions.

Table 2. Group Intercorrelations of Writing Variables (n=98)

Variable	1	2	3	4	5	6	7	8
1. Hours Typing	--	.17	.15	.11	.20	.31**	.12	.07
2. Keyboarding	.16	--	-.03	-.01	.18	.11	-.05	-.03
3. WP Quality	-.01	.36	--	.67**	.43**	.37**	.51**	.48**
4. WP Length	.10	.26	.76**	--	.37**	.49**	.44**	.55**
5. HW Quality	.07	-.12	.38*	.30	--	.54**	.36**	.17
6. HW length	-.04	.01	.62**	.54**	.59**	--	.27*	.14
7. Writing Fluency	-.32	-.02	.48**	.28	.26	.43*	--	.45**
8. NWT Task	-.09	.19	.53**	.63**	.26	.37*	.38*	--

Note. Correlations above the diagonal were observed in the NLD group; those below the diagonal were observed in the LD group. * $p < .05$. ** $p < .01$.

Table 3. Collapsed Group Intercorrelations of Writing Variables (n=98)

Variable	1	2	3	4	5	6	7	8
1. Hours Typing	--	.17	.12	.11	.16	.22*	.01	.02
2. Keyboarding		--	.07	.05	.08	.08	-.04	.04
3. WP Quality			--	.70**	.43**	.43**	.49**	.50**
4. WP Length				--	.37**	.51**	.39**	.57**
5. HW Quality					--	.56**	.33**	.21*
6. HW Length						--	.31**	.21*
7. Writing Fluency							--	.43**
8. NWT Task								--

Note. Correlations were collapsed across groups. * $p < .05$. ** $p < .01$.

Regression analyses (simultaneous entry of variables) were conducted to determine the extent to which variables might predict quality scores on the essays for both formats (WP and HW). Analyses were run using group (NLD and LD), length, WJ III Writing Fluency raw score, and number writing (# of letters per minute) as predictor variables. For quality scores in the WP format, results produced $R^2 = .55$, $F(4, 98) = 28.36$, $p < .001$ for the overall model. WP length contributed 18.5% of the variance and writing fluency contributed 5% of the variance in the model. No other variables made significant contributions. For quality scores in the HW format, results produced $R^2 = .35$, $F(4,98) = 12.58$, $p < .001$. HW length contributed most to this model by explaining 22.1% of the variance.

2.6 Group Comparisons

Table 4. Means and Standard Deviations of Dependent Measures (Standard Deviations are in Parenthesis).

Group	NLD (<i>n</i> =68)	LD (<i>n</i> =30)	t-test	p-value	<i>d</i>
Writing Fluency	28.1(.6)	28.1(.8)	.01	.99	.00
Calculations	127.9(22.9)	117.2(26.1)	1.96	.06	.04
NWT	187.5(46.6)	181.6 (49.0)	.57	.57	.00
WP Quality ^a	4.01 (1.2)	3.6 (.96)			
HW Quality	3.40 (.87)	3.1 (.90)			
WP Length ^b	265.9 (87.9)	234.3 (75.2)			
HW Length	180.7 (40.19)	176.1 (38.47)			

^a There was no significant group by format interaction for quality; there was a main effect showing greater quality in the WP condition.

The sample size was obtained through the use of G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007). Using an effect size (ES) of 0.70, alpha of 0.05, and a power value of 0.80 with two groups, the total number of participants required for sufficient results using a one-tailed analysis of variance was $N = 52$. Group t-test comparisons were conducted on several performance measures including the Calculations Subtest, NWT, and Writing Fluency Subtest (see Table 4). The comparison on Calculations score just failed to reach significance ($p = .06$). Even though the difference is close to significant, the effect size is minimal ($d = .04$). Certainly it is plausible that some students with LD experience difficulties in mathematics. The groups were found to exhibit similar typing speed abilities ($p = .57$), suggesting that typing speed is not a confounding factor in further analyses. In order to explore differences between NLD and LD groups in writing skills, the groups were compared on the Writing Fluency task. No significant group differences were found ($p = .99$). This was an unexpected finding, suggesting no differences in the speed of writing between the two groups.

2.7 Analysis of Primary Hypotheses

Order

A 2x2 analysis of variance (ANOVA) was used to examine order effects for format (HW and WP) and WIAT-II essay prompt (A and B) to determine if the order in which formats and prompts were administered had a significant effect on results. The analysis revealed no significant interaction or main effects for order (all p-values > .10). T-tests were then conducted to determine if there were differences between male and female responses. There were no significant sex differences in quality in the WP format (males $M = 3.73$,

$SD = 1.16$, females $M = 4.06$, $SD = 1.09$, $t(97) = -1.50$, $p = .15$) or HW format (males $M = 3.23$, $SD = .98$, females $M = 3.41$, $SD = .75$, $t(97) = -1.02$, $p = .31$). However, there were significant sex differences in regard to length with females writing significantly longer essays than males in both WP (males $M = 237.81$, $SD = 71.98$, females $M = 276.81$, $SD = 94.27$, $t(97) = -2.33$, $p < .05$) and HW formats (males $M = 171.40$, $SD = 36.23$, females $M = 188.02$, $SD = 41.56$; $t(97) = -2.13$, $p < .05$). Sex was included as a variable in subsequent analyses of essay length.

Quality Scores

According to the Interaction Hypothesis, it was predicted that a significant interaction would be expected between group (LD and NLD) and format (HW and WP). Students with LD were expected to improve significantly more than NLD students from the HW (standard) condition to the WP (accommodated) condition in respect to quality of essays. Table 4 contains the means for each group for quality score in both formats. A 2 Group (LD and NLD) X 2 Format (HW and WP) mixed measures ANOVA revealed that there was no significant interaction between group and format with respect to quality, $F(1,96) = .24$, $p = .63$. This shows that the students with LD, compared to the NLD students, did not increase significantly from HW to WP formats in regard to quality scores. In other words, they did not experience a “differential boost.” A significant main effect for format was found for the quality score, with increased performance in WP condition ($M = 3.90$, $SD = 1.16$) compared to the HW condition ($M = 3.32$, $SD = .88$). Both groups performed better in WP condition, $F(1,96) = 21.23$, $p < .001$, $d = .18$. There were no significant group differences for either condition, although the NLD group tended to have higher mean scores for WP ($d = .38$) and HW ($d = .34$).

Length

Based on the interaction hypothesis, it was predicted that a second significant interaction would be expected between group and format. Students with LD were expected to improve significantly more than NLD students from the HW condition to the WP condition in respect to length (total word count). Table 4 contains the means for each group for length in both formats. Since sex was found to be significant in regard to length in both HW and WP formats, this variable was initially included in the analysis. A three-way mixed measures ANOVA conducted on group X sex X format revealed no significant three-way interaction among the variables, $F(1,95) = .11$, $p = .74$. The interaction hypothesis was then analyzed using group and format variables. This analysis revealed no significant two-way interaction between group and format with respect to length, $F(1,95) = 2.86$, $p = .09$. This shows that students with LD compared to the NLD students did not increase significantly from HW to WP formats in regard to length. The main effect for length on the two formats was significant, $F(1,97) = 81.15$, $p < .001$, $d = .46$, with the WP essays ($M = 256.32$, $SD = 85.15$) showing more total words than the HW essays ($M = 179.29$, $SD = 39.54$) for both LD and NLD groups. There were no significant main effects for group in either format condition, although the mean difference in the WP condition had an effect size of $d = .39$.

2.8 Exploratory Hypothesis

Time

An additional exploratory analysis was conducted for time as a dependent measure. It was hypothesized that total amount of time spent on essays would be greater for the LD group than the NLD group. In other words, would the percent of LD students who used the entire 10-minute time limit be greater than the percentage of NLD students? It should be noted there was a ceiling on time at 10-minutes in order to create a time-sensitive task. The percentage of participants who used the entire 10-minute time limit to complete the essays differed by group. In the WP condition, 93% of the NLD group and 67% of the LD group used the entire time. In the HW condition, 88% of the NLD group and 77% of the LD group used the entire time. Chi-square analyses were conducted between groups (NLD and LD) and completion time in both formats (WP and HW). In the WP condition, significant differences were found between groups with a greater percentage of students in the NLD than LD group using the full 10-minutes, $\chi^2(1, N = 98) = 27.02, p < .01$. In the HW condition, no significant differences were found between groups $\chi^2(1, N = 98) = 19.92, p = .18$. The completion difference was surprising given that students with LD receive extended time, yet tended to finish sooner.

In addition, student preference for WP vs. HW and ease of response mode (WP vs. HW) were analyzed. Seventy-four percent of students in both groups reported that they preferred to use a WP to HW and 61% in both groups reported that they found typing on a WP easier than HW.

3. Discussion

The hypothesis that a significant interaction would occur between group (NLD and LD) and format (HW and WP) with respect to quality and length of essays was not supported. Both groups improved their performance in the WP compared to HW condition, thus there was no evidence of an interaction or differential boost for students with LD. This is consistent with similar studies that do not show support for the interaction hypothesis regarding extended time for students with disabilities (Alster, 1997; Fuchs et al., 2000a; Lewandowski, Lovett, & Rogers, 2008; Lewandowski, Lovett, Parolin, Gordon, and Coddling, 2007; Weaver, 2001). Such studies show extended time as a benefit to students with and without a disability, and therefore do not support the specificity criterion (Fuchs et al., 2000b; Phillips, 1994; Sireci et al., 2005; Zuriff, 2000). According to Phillips (1994), accommodations are specific when they significantly assist students with disabilities in performing a task or test, but have little or no effect on nondisabled students. The present study showed that word processors, as accommodations, are nonspecific. As the case with extended time, word processors confer a performance advantage for all students regardless of disability status. Both groups improved equally in writing quality and length when using a WP, therefore showing no differential boost for the LD group.

The performance advantage of a WP for all students is a novel and important finding. This study shows that all students benefited significantly in the WP format. Such an outcome can be generalized to classroom and high stakes exams. If students are able to take writing exams on a WP, they may be able to perform at a higher qualitative level (organization, theme development, and vocabulary). This performance advantage was not apparent in previous research exploring such variables. Most research comparing formats (WP and HW) found that the HW condition received higher quality scores than the WP condition as a result of rater bias (Arnold et al., 1990; Bridgeman & Cooper, 1998; McCann, Eastment & Pickering, 2002; Powers et al., 1994; Russell & Wei, 2004). It is important to note that in almost all previous research, essays were scored in the format in which they were originally written (HW scored in handwritten format and WP scored in word processed format), subjecting such studies to rater bias. Previous studies that did transcribe all essays to a single format did not have raters score the essays blindly; rather, raters were aware of the original format of the essays and tended to favor HW over WP (Hollenbeck et al., 1999; Russell & Wei, 2004). In the current study, all HW essays were transcribed to WP format and scored blindly by raters (raters did not know if they were reading a WP essay or HW essay), which essentially eliminated rater bias. It might be important to note, if researchers continue to study both formats, they should consider a design that converts written essays to typed essays for scoring.

The performance advantage of a WP for students with LD and NLD was not just limited to quality scores, but was also apparent in length of essays. As with quality score, essay length (total words written) increased for both groups from the HW to WP condition, with the LD group increasing by an average of 58 words and the NLD group an average of 85 words. In fact, all students, regardless of disability status or sex, wrote longer essays in the WP condition. This finding is supported by previous research that found students wrote longer essays on a WP than by hand (Gregg et al., 2007; Russell et al., 1999; Truell, Alexander & Davis, 2004). Therefore, a WP appears to benefit all students in writing production. In addition, results indicate that essay length correlated highly with quality score. The relationship between the amount of words written and quality of the essays for the entire sample was $r = .70$ in the WP, and $r = .56$ in the HW conditions. This finding is consistent with previous research that shows a high correlation between amount of words written and quality scores (Espin et al., 2005; Ewing et al., 2005; Gregg et al., 2007; Kobrin & Kimmel, 2006). Interestingly, the exploratory hypothesis, stating that total amount of time spent on essays would be greater for the LD group than the NLD group, was not supported. In fact, in the WP condition, significant differences were found between groups with a greater percentage of students in the NLD than LD group using the full 10-minutes. It is possible that the LD group finished writing their essays too soon. Previous research has shown support for this hypothesis with results indicating that students with LD have a difficult time with planning and executing essays (Graham & Harris, 2003; Graham, Harris, & Troia, 1998). One would expect the LD group would spend more time writing essays given

that such students often receive extended time for exams and assignments. These findings have implications for the validity of extended time for students with LD.

However, such findings, when considered in terms of classroom and standardized tests, may actually raise skill level, save time, and be more functional to grade than handwritten tests. Therefore, rather than viewed as a test accommodation, perhaps WP should be considered as a standard writing format whenever possible. Another benefit of the WP format is the possibility of spelling and grammar check, as well as ease of editing. If the writing test attempts to measure quality of ideas, organization, thematic cohesion, and vocabulary, then features like spell check and grammar editing could improve the presentation and readability of the writing, making it easier for a grader to evaluate the relevant dimensions of the task. It would seem that the WP format offers greater advantages than the HW format, while being fair for males and females, and students with and without disabilities. Perhaps it is time to allow students to write with today's technology instead of the old fashioned paper and pencil format.

3.1 Limitations

Several limitations should be noted when considering the generalizability of the findings. First, although students with LD selected for this study show documented professional diagnoses that have been reviewed and approved by an accredited university, this study showed no significant differences between the LD and NLD groups on the writing fluency task or quality and length of essays, although there were small effect size differences (.34 - .39) in favor of the NLD group. This begs the question of whether or not these groups are really different from one another in ability and performance. It could be that students in the LD group were not properly diagnosed in accordance with the DSM-IV criteria, particularly with regard to significant impairments in academic skill levels. After all, how impaired could students with LD be who are, as a group, attaining a 3.0 average at a major research university? Sparks and Lovett (2009) reported that in a large sample of college students with LD diagnoses, most LD students actually showed average (not impaired) achievement test scores. It would be helpful if college disability administrators insured that LD students had impairment in specific academic skills and granted accommodations especially for a given impairment.

It is possible that the LD diagnoses in this study were valid, yet the measures used in this study were not sensitive enough to detect group differences. The only evidence of a difference in group performance was GPA. It is possible that differences in GPA could have resulted from any number of reasons including lower academic expectations by the LD group or the effects of a heavy academic workload. It would have been ideal to have a large sample of LD students that actually demonstrated significant impairment in a particular aspect of writing, but we could not find such a group in a university of 12,000 students.

Despite the questions concerning diagnosis, the LD group in this study can at least be considered as ecologically valid. That is, students in this study received external

verification of their disability status in line with DSM IV-TR criteria, and were found to meet qualifications for test accommodations under the Americans with Disabilities Act (1990). Almost all students were approved for a WP as an accommodation, and reasons for this varied on their accommodation plans (i.e., poor handwriting, spelling, organization, and weak language processing). In this light, the study findings may only generalize to current college students in the United States who qualify to receive test accommodations based on an LD diagnosis.

3.2 Future Research

Future research in this area would be better served if LD diagnoses can be verified through direct testing, and LD participants actually demonstrate impairment in areas for which they receive accommodations (i.e., writing fluency or spelling). Further, we need research that is international or cross-cultural in design so that we can better understand the procedures and policies various countries use for verifying disabilities and granting test accommodations. Of course, there are many variables that can be explored with regard to computerized writing beyond length and quality measures. The value of a word processor for spelling, grammar, vocabulary, revising, and so on should be a focus of research for students with and without disabilities.

In conclusion, and despite noted limitations, this study suggests that a WP was beneficial for students with and without LD. Use of the computer conferred a performance advantage on essay writing tasks in the areas of quality and length. It is also preferred by students and found to be easier to use. Since word processors have become such an important part of our culture, it may be valid for all students to use word processors for written essays and exams. Future research could explore the use of computers on classroom-wide exams and high-stakes tests across grade levels and types of students.

The performance advantage of a WP for all students is a novel and important finding. As mentioned previously, the outcomes of this study can be generalized to classroom and high stakes exams. Such findings, may actually raise skill level, save time, and be more functional to grade than handwritten tests. It will not be long before this writing format is the rule and not the exception.

References

- Alster, E.H. (1997). The effects of extended time on algebra test scores for college students with and without learning disabilities. *Journal of Learning Disabilities*, 30, 222-227.
- Americans with Disabilities Act of 1990, P.L. 101-336, 42 U.S. C.A. 12, 101-12, 213 (West Supp. 1991).
- Arnold, V., Legas, J., Obler, S., Pacheco, M.A., Russell, C., & Umbdenstock, L. (1990). *Do students get higher scores on their word-processed papers? A study of bias in scoring hand-written vs. word-processed papers*. Whittier, CA: Rio Hondo College. (ERIC Document Reproduction Service No. ED 345818)

- Bangert-Drowns, R. L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research, 63*, 69-93.
- Berninger, V.W., Abbott, R.D., Augsburger, A., & Garcia, N. (2009). Comparison of pen and keyboard transcription modes in children with and without learning disabilities. *Learning Disability Quarterly, 33*, 123-141.
- Bolt, S. E., & Thurlow, M. L. (2004). Five of the most frequently allowed testing accommodations in state policy. *Remedial and Special Education, 25*, 141-152.
- Bridgeman, B., & Cooper, P. (1998). *Comparability of scores on word-processed and handwritten essays on the Graduate Management Admission Test*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Espin, C.A., De La Paz, S., Sclerka, B.J., & Roelofs, L. (2005). The relationship between curriculum-based measures in written expression and quality and completeness of expository writing for middle school students. *The Journal of Special Education, 38*(4), 208-217.
- Ewing, M., Huff, K., Andrews, M., & King, K. (2005). *Assessing the reliability of skills measured by the SAT* (College Board RN-24). Princeton, NJ: Educational Testing Service.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.
- Fuchs, L.S., Fuchs, D., Eaton, S.B., Hamlett, C., Binkley, E., & Crouch, R. (2000a). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children, 67*(1), 67-81.
- Fuchs, L.S., Fuchs, D., Eaton, S.B., Hamlett, C.L., & Karns, K.M. (2000b). Supplementing teacher judgments of mathematical tests accommodations with objective data sources. *School Psychology Review, 29*(1), 65-85.
- Gordon, M., Lewandowski, L., & Keiser, S. (1999). The LD label for relatively well-functioning students: A critical analysis. *Journal of Learning Disabilities, 32*, 485-490.
- Graham, S., & Harris, K.R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In H.L. Swanson, K.R. Harris, & S. Graham (Eds.), *Handbook of Learning Disabilities* (pp. 323-344). New York: Guilford.
- Graham, S., Harris, K., & Hebert, M. (2011). *Informing writing: The benefits of formative assessment. A Carnegie Corporation Time to Act report*. Washington, DC: Alliance for Excellent Education.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*, 445-476.
- Graham, S., Harris, K.R., & Troia, G.A. (2000). Self-regulated strategy development revisited: Teaching writing strategies to struggling writers. *Topics in Language Disorders, 20*(4), 1-14.
- Gregg, N., Coleman, C., Davis, M., & Chalk, J.C. (2007). Timed essay writing: Implications for high-stakes tests. *Journal of Learning Disabilities, 40*(4), 306-318.
- Henderson, C. (2001). *College freshman with disabilities: a biennial statistical profile*. Washington, D.C.: American Council on Education.
- Hollenbeck, K., Tindal, G., Stieber, S., & Harris, M. (1999). *Handwritten versus word processed statewide compositions: Do judges rate them differently*. Eugene, OR: University of Oregon. Individuals with Disabilities Education Act of 1991, 20 U.S.C. 1400 et seq. (1991).
- Horkay, N., Elliot Bennett, R., Allen, N., Kaplan, B., & Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *The Journal of Technology, Learning, and Assessment, 5*(2). Retrieved September 8, 2007 from <http://www.jtla.org>.
- Hughes, C.A., & Osgood Smith, J. (1990). Cognitive and academic performance of college students with learning disabilities: A synthesis of the literature. *Learning Disability Quarterly, 13*(1), 66-79.
- Kobrin, J.L., & Kimmel, E.W. (2006). *Information on the writing section of the SAT Reasoning Test* (College Board RN-25). Princeton, NJ: Educational Testing Service.

- Lam, F.S., & Pennington, M.C. (1995). The computer vs. the Pen: A comparative study of word processing in a Hong Kong secondary classroom. *Computer Assisted Language Learning*, 8, 75-92.
- Lewandowski, L., Lovett, B. J., Parolin, R., Gordon, M., & Coddington, R. S. (2007). Extended time accommodations and the mathematics performance of students with and without ADHD. *Journal of Psychoeducational Assessment*, 25, 17-28. doi: 10.1177/0734282906291961
- Lewandowski, L., Lovett, B. J., & Rogers, C. L. (2008). Extended time as a testing accommodation for students with reading disabilities: Does a rising tide lift all ships? *Journal of Psychoeducational Assessment*.
- Li, H., & Hamel, C.M. (2003). Writing issues in college students with learning disabilities: A synthesis of the literature from 1990 to 2000. *Learning Disability Quarterly*, 26, 29-46.
- Lovett, B. J. (2011). Extended time testing accommodations for students with disabilities: Answers to five fundamental questions. *Communique*, 39, 14-15.
- Lovett, B. J., & Lewandowski, L. J. (2006). Gifted students with learning disabilities: Who are they? *Journal of Learning Disabilities*, 39, 515-527.
- Macann, R., Eastment, B., & Pickering, S. (2002). Responding to free response examination questions: computer versus pen and paper. *British Journal of Educational Technology*, 33(2), 173-188.
- MacArthur, C.A. (1996). Using technology to enhance the writing process of students with learning disabilities. *Journal of Learning Disabilities*, 29(4), 344-354.
- MacArthur, C.A., & Graham, S. (1987). Learning disabled students' composing under three methods of text production: Handwriting, word processing, and dictation. *The Journal of Special Education*, 21(3), 22-42.
- Maccini, P., Gagnon, J.C., & Hughes, C. (2002). Technology-based practices for secondary students with learning disabilities. *Learning Disability Quarterly*, 25, 247-261.
- Mogey, N., Paterson, J., Burk, J., & Purcell, M. (2010). Typing compared with handwriting for essay examinations at University: Letting the students choose. *Research in Learning Technology*, 18, 29-47.
- Morphy, P., & Graham, S. (2012). Word processing programs and weaker writers/readers: A meta-analysis of research findings. *Reading and Writing*, 25(3), 641-678.
- Phillips, S.E. (1994). High-Stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education*, 7 (2), 93-120.
- Powers, D.E., Fowles, M.E., 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.
- Ranssen, J. D., & Parks, G. S. (2005). Test accommodations for postsecondary students: The quandary resulting from the ADA's disability definition. *Psychology, Public Policy, and Law*, 11(1), 83-108.
- Russell, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives*, 7(20). Retrieved January 28, 2006 from <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). Retrieved January 28, 2006 from <http://epaa.asu.edu/epaa/v5n3.html>.
- Russell, M. & Wei, T. (2004). The influence of computer-print on rater scores. *Practical Assessment, Research & Evaluation*, 9(10). Retrieved February 1, 2006 from <http://PAREonline.net/getvn.asp?v=9&n=10>.
- Sireci, S.G., Scarpatti, S., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75, 457-490. doi: 10.3102/00346543075004457

- Sparks, R., & Lovett, B. J. (2009). College students with learning disability diagnoses: Who are they and how do they perform. *Journal of Learning Disabilities, 42*, 494-510. doi: 10.1177/0022219409338746
- Truell, A. D., Alexander, M. W., & Davis, R. E. (2004). Comparing postsecondary marketing student performance on computer-based and handwritten essay tests. *Journal of Career and Technical Education, 20*, 69-78.
- U.S. Department of Education, National Center for Educational Statistics. (2011). *Digest of Education Statistics, 2010* (2011-015), Chapter 2.
- Van Waes, L., Leijten, M., & Quinlan, T. (2010). Reading during sentence composing and error correction: A multilevel Analysis of the influences of task complexity. *Reading and Writing: An Interdisciplinary Journal, 23*, 803-834. doi: 10.1007/s11145-009-9190-x
- Van Waes, L., & Schellens, P. J., (2003). Writing profiles: the effect of writing mode on pausing and revision patterns of experienced writers. *Journal of Pragmatics, 35*, 829-853. doi: 10.1016/S0378-2166(02)00121-2
- Weaver, S.M. (1993). *The validity of the use of extended and untimed testing for postsecondary students with learning disabilities*. Unpublished doctoral dissertation, University of Toronto, Canada.
- Wechsler, D. (2002). Wechsler Individual Achievement Test (2nd Ed.). San Antonio, TX: Psychological Corporation.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson Tests of Achievement*. (3rd ed.). Itasca, IL: Riverside Publishing.
- Zuriff, G.E. (2000). Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Applied Measurement in Education, 13* (1), 99-117.