

## THE EFFECTS OF EXTENDED TIME ON WRITING PERFORMANCE

**Lauren D. Goegan**

*University of Alberta, Canada*

**Gina L. Harrison**

*University of Victoria, Canada*

---

*The effects of extended time on the writing performance of university students with learning disabilities (LD) was examined. Thirty-eight students (19 LD; 19 non-LD) completed a collection of cognitive, linguistic, and literacy measures, and wrote essays under regular and extended time conditions. Limited evidence was found to support the accommodation of extended time to improved writing quality indices. Compared to the students without LD, the students with LD demonstrated increased writing fluency with extended time, but the organizational and thematic quality of their essays continued to be lower. The results are discussed in relation to current theory and research on the persistence of writing disabilities into adulthood, and the provision of empirically supported accommodations for students with LD in higher education.*

---

**Keywords: Learning Disabilities University, Writing, Accommodations, Extended Time**

### INTRODUCTION

The overarching premise in accommodating students with learning disabilities (LD) is to ensure that their performance outcomes reflect the same attributes, abilities, and knowledge as their non-LD peers (Gregg & Nelson, 2010; Lovett, 2010). Although specific statistical estimates vary by country, jurisdiction, and institution, extended time has been reported to be the most common accommodation requested and provided to students with LD at the postsecondary level (Lovett, 2010; Ofiesh 2000; Ofiesh, Hughes, & Scott, 2004; Zuriff, 2000). Despite its reportedly high prevalence, limited research has been conducted on the efficacy of extended time on test performance for all students, not only students with LD (e.g., Sereci, Scarpati, & Li, 2005; Zuriff, 2000). Of the studies available, extended time is defined, on average, as time and one-half to double the standard allotted time (Ofiesh & Hughes, 2002). Furthermore, in the only meta-analysis conducted with few studies available, Gregg and Nelson (2012) reported that for adolescents transitioning to higher education, students without LD outperformed students with LD irrespective of standard or extended time conditions and of the academic skills area assessed (e.g., reading/writing or math). Weis, Dean, and Osborne (2014) also reported that nearly 90% of clinicians (based on 350 LD diagnoses) recommended extra time in their assessment reports for students in higher education across subject areas and individual cognitive and achievement profiles, despite a lack of clinical and research-informed evidence of its efficacy.

Since clinicians have a tendency to recommend extra time as a “one size fits all” for students with LD (e.g., Weiss et al., 2014), we aimed to study whether a

---

\*Please send correspondence to: Gina L. Harrison, PhD, Department of Educational Psychology and Leadership Studies, University of Victoria, 3010 STN CSC, Victoria, B.C. V8W 2Y2, Canada, Phone: 1-250-721-7783, Email: harrison@uvic.ca.

sample of students with LD would show improved writing performance with extra time drawing on two theories. First, the *interaction hypothesis* (Sereci, Scarpata, & Li, 2005), also referred to as the *maximum potential thesis* (Zuriff, 2000), contends that only students with LD benefit from extended time; students without LD will show no improvement in performance as they are already performing at their *maximum potential* with regular time (Lovett, 2010; Stretch & Osborne, 2005; Zuriff, 2000). Evidence across two sources is needed to test this theory: (1) Students with LD must show significantly improved scores when given additional time compared to their scores under regular time, and (2) students without LD must not score significantly higher when provided with extended time compared to regular time (Sereci et al., 2005) denoting an interaction between LD status and time. A second less conservative approach involves evaluating whether or not students with LD experience a *differential boost* in their performance under extended time conditions (Fuchs & Fuchs, 2001). That is, all students benefit from extended time to some degree, but that there is a differential boost in scores for students with LD who improve their performance from regular to extended time conditions significantly more than students without LD (Fuchs & Fuchs, 2001; Lewandowski, Lovett, & Rogers, 2008; Sireci, et al., 2005). There is no consensus in the literature, however, on whether this boost in performance must be “statistically” significant under extended time conditions.

Evaluating the efficacy of extended time against either of these theories requires that the same measure be completed by students with and without LD under both regular and extended times. Only two studies to date have employed a repeated measures design and both studies provide evidence consistent with a *differential boost* in performance for the students with LD. In the earlier study, college students with LD and their non-LD peers completed an algebra test under regular time (12 minutes) and extended time (unlimited) conditions (Alster, 1997). Students in the LD group scored significantly lower than the non-LD group under the regular time conditions, but no differences were detected between the groups with extended time as both groups significantly improved their performance. Consistent with the *differential boost* hypothesis, the authors also reported that the performance gain for the LD group was of a significantly greater magnitude than for the non-LD group.

Similarly, Lesaux and colleagues (2006) examined the reading comprehension skills of university students with reading disabilities in relation to average and above average readers on a reading comprehension test. A norm-referenced reading measure was administered under regular time (20 minutes) and an untimed condition (up to 40 minutes). Two equivalent forms of the measure were counterbalanced across groups and time conditions. Scores on the reading comprehension task were compared between the time conditions and group status. The students with LD benefited significantly from extra time, while their non-LD peers showed some improvement with extended time, the scores attained were not significantly different than under regular time. Even with extended time, however, the students with more severe reading disabilities still performed significantly lower on the reading comprehension measure than students without LD. Consistent with Gregg and Nelson's (2012) conclusions, extended time does not eradicate the academic skills deficits that endure into adulthood for students with LD in higher education.

To date, no studies have examined the efficacy of extended time on the writing performance for students with LD in relation to either theory. This lack of research is surprising given the persistent difficulties in writing experienced by adolescents and adults with LD (e.g., Connelly, Dockrell, & Barnett, 2005; Connelly, Campbell, MacLean, & Barnes, 2006; Harrison, 2009). Writing is also the academic skill that is reportedly the most challenging for students with LD in higher education (e.g., Hatcher et al., 2002). The provision of extended time ostensibly allows students the opportunity to demonstrate their knowledge on a topic (Ofiesh, Hughes & Scott, 2004) without reducing the quality of their written responses due to limitations in speed and efficiency (Lindstrom, 2007). Writing presents numerous challenges to students with LD, including mechanical (i.e., lower order skills) and compositional difficulties (i.e., higher order skills; Berninger, 1999; Harrison, 2009; Lindstrom, 2007) that impact the quality of a text at its microstructure (sentence level) and macrostructure (total text). Processing speed deficits may be related to a failure to automatize basic skills such as handwriting fluency and spelling (Ofiesh, 2000). This lack of automatization may be further impacted by a limited capacity of working memory (WM) and other executive functions integral to coordinating a writing task (Berninger & Amtmann, 2003). In a study with college students, Gregg, Coleman, Davis and Chalk (2007) reported that indices of lexical complexity (i.e., number of different words, number of words with two or more syllables, and number of letters per word), text length (i.e., total word count, also an index of writing fluency), spelling, and handwriting explained more variance in essay quality on a timed writing task for students with LD (i.e., dyslexia) compared to their non-LD peers. We included the same scoring for lexical quality in the present study to see whether there would be any differences between groups under regular and extended time conditions.

The quantity (i.e., word count) and quality across both lower order skills (i.e., spelling, mechanics) and higher order skills (i.e., theme development, organization, vocabulary) of writing produced by university students with a history of LD has been reported as significantly lower compared to their non-LD peers (Harrison & Beres, 2007). In a literature synthesis, Li and Hamel (2003) investigated the characteristics and error patterns of the writing of students in higher education with LD. Of only seven articles that were included, Li and Hamel (2003) highlighted two categories of errors: mechanical (e.g., spelling, punctuation, capitalization) and content (e.g., planning, text coherence, organization). Gregg, Hoy, Alexander and Hayes (1991) noted that writers with LD experience processing difficulties impeding their ability to acquire and/or adhere to syntactic and mechanical conventions when they write. Error types across mechanical components (e.g., spelling, punctuation, capitalization) as well as errors across the higher level aspects of text organization and coherence can reflect difficulties with written syntax and grammar (Li & Hamel, 2003). The writing skills of students with LD reflects a lack of automaticity in lower-order aspects of writing (e.g., handwriting fluency and spelling) constraining working memory and cognitive control resources available to perform the higher-order aspects of writing (e.g., generating ideas, organizing texts, etc.; Connelly et al., 2005; Connelly et al., 2006; Harrison, 2009). Writing difficulties can also be intractable in higher education where curriculum demands and instructor expectations intensify, compounding the challenges students experience with writing (Li & Hamel 2003; Lindstrom, 2007).

Defining improvement to writing performance under regular and extended time conditions compared to performance in math and reading comprehension is more complex. With respect to reading comprehension and math, improved performance has been defined as increased accuracy in the number of items completed (e.g. Alster, 1997; Lesaux et al., 2006). Evaluating improved writing performance is complicated by the fact that writing is multifaceted and that evaluations of text quality involve multiple components usually across a continuum of ratings. The present study is the first to employ a repeated measures design to examine writing performance with and without extended time across lower-order (mechanics: spelling, punctuation, and capitalization and word count as an index of text length) and higher order (theme development, organization, vocabulary and lexical quality) aspects of writing performance as indices of writing quality in students with LD in higher education. We evaluated the effects in relation to the two theories previously used in the literature: the *differential boost* and *interaction* hypotheses. The following questions were investigated: (1) Does extended time improve (i.e., increase) the overall scores obtained on written essays when evaluated across the aforementioned lower-order and higher-order components? (2) If writing performance does improve with extra time, then which aspects (i.e., lower- or higher-order) improve? (3) If there are increases in writing performance with extended time, do students without LD also benefit? (4) Is there evidence of a *differential boost* or an *interaction* between LD status and time in the writing performance of students with LD?

## METHOD

### *Participants*

A total of 38 students (19 LD, 19 non-LD) participated from a mid-size (about 22,000 students) Canadian university in a central urban area. All students were undergraduates with most (78%) in their first 3 years of 4-year programs across a variety of disciplines. Students with LD (11 men, 8 women; Mean age = 24.22, SD = 5.95) had a formal diagnosis (based on DSM-IV-TR criteria) and low writing achievement, were registered with the campus disability resource centre, and were provided with the accommodation of extended time for tests. Confirmation of diagnosis, medical, and educational history was obtained through a student self-report questionnaire considered a valid method for determining adults' literacy-based learning difficulties in research (McGonnell, Parilla & Deacon, 2007). Students had been diagnosed with LD for about 10 years (mean = 10.78 years) and had a history of special education during elementary and high school. Students without LD (14 women, 5 men; Mean age = 21.99, SD = 2.54) were recruited through the research participant pool in the Psychology Department and had no history of learning difficulties or special education. No information on ethnic background was collected, and participants were excluded from either group if they had a primary language other than English and if they had a co-occurring medical or mental health condition.

### *Measures*

A collection of cognitive, linguistic, and literacy measures known to be associated with literacy-based LD were administered to participants to assess individual differences in these hallmark skills and processes between the two groups. Scaled

scores ( $M = 10$ ,  $SD = 3$ ) or standard scores ( $M = 100$ ,  $SD = 15$ ) were recorded and used in the analyses for norm-referenced measures.

**Verbal Short-term and Verbal Working Memory.** The Digit-Span Forward (DSF) and the Digit Span Backward (DSB) tasks from the Wechsler Adult Intelligence Scale 3<sup>rd</sup> Edition (WAIS-III; Wechsler, 1997) were administered to assess verbal short-term and verbal working memory (WM), respectively. DSF required the student to repeat increasingly longer number sequences verbatim and DSB required the student to repeat increasingly longer sequences of digits in reverse order. The tasks were administered and scored according to the procedures described in the WAIS-III manual, which reports internal consistency estimates as .88 and .82 for DSB and DSF, respectively. Raw scores were recorded for each task, as separate scaled score conversions were not available in the manual.

**Rapid Automatized Naming.** The RAN-Digits and RAN-Letters subtests from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Raschotte, 1999) were used to assess rapid naming. Participants were required to quickly and accurately name a random array of digits (0-9) or letters. Test-retest reliability as reported in the manual is .90 and .86 for RAN-Digits and RAN-Letters, respectively. The measure was administered and scored according to the standardized procedures in the manual and scaled scores were recorded.

**Expressive Vocabulary.** The vocabulary subtest of the Wechsler Adult Intelligence Scale-3<sup>rd</sup> Edition (WAIS-III, Wechsler, 1997) was administered. Participants were required to orally define words that were shown to them and orally presented by the examiner. The manual reports a strong internal consistency estimate of .88 for the age range of our participants, and the administration and scoring followed the procedures described in the test manual.

**Word Reading and Decoding Fluency.** The two subtests from the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999) were administered to assess word reading and decoding fluency. The word reading and non-word reading tasks required participants to read aloud as many words (word reading) and pseudowords (nonword reading) as possible in 45 seconds. Test-retest reliability is .82 and .89 for adults in the standardization sample for the word reading and non-word reading subtests, respectively.

**Spelling.** The Spelling subtest from the Wide Range Achievement Test-Third Edition (WRAT-3; Wilkinson, 1993), blue form was used to assess spelling in isolation. This task is a spelling dictation task, administered and scored according to the procedures described in the manual. Internal consistency for the age range of our participants is .82 as reported in the manual.

**Handwriting Fluency.** Handwriting fluency was assessed using an experimental measure developed by Berninger, Mizokawa and Bragg (1991). This task requires participants to write the letters of the alphabet in order, in lowercase letters, as quickly as they can in one minute. Raw scores were calculated by two research assistants for each participant, and disagreement between the number of legible letters written were resolved through discussion. Total number of legible letters were recorded.

**Essay Writing.** Students completed the essay composition subtest from the Wechsler Individual Achievement Test-2<sup>nd</sup> Edition (WIAT-II; The Psychological

Corporation; 2002) which has the advantage of two prompts (A and B) that we counterbalanced across participants, and across time conditions (regular or extended). Essays are evaluated across lower-order and higher-order aspects consistent with contemporary componential models of writing (e.g., Berninger & Amtmann, 2003). This task is also considered an ecologically valid writing measure as participants are required to write a persuasive essay, similar to the academic writing demands of their coursework (Connelly et al., 2006). The task was administered according to standardized instructions, with the exception that the extended time condition was twice as long (30 minutes) as the regular time condition (15 minutes). We used the double time metric for extended time to align with the metric used in other studies. Ofesh & Hughes (2002), found that when students were provided with unlimited time, most students used double time. Furthermore, while Leaux and colleagues (2006) referred to their extended time condition as “untimed,” there was actually a limit of double time. Additional information about administration and the prompts for the essays can be found in the WIAT-II manual (Psychological Corporation, 2002).

**Scoring.** Essays were transcribed via a word processor maintaining errors in spelling and punctuation so that any bias in essay quality due to poor handwriting was eliminated (Graham & Weintraub, 1996). Each essay was transcribed by a research assistant, and then confirmed by a second research assistant. Word count (text length) was established based on the built-in word counter in the word processing program.

Essays were scored across four categories as described in the test manual: mechanics, organization, theme development and vocabulary. Mechanics (max. score 9) included spelling and punctuation (e.g. capitalization, comma use) errors. Organization (max. score 17) assessed elements such as sentence structure and use of linking expressions. Theme development (max. score 8) examined the reasons given for students’ positions and support for their arguments on the question posed. Vocabulary (max. score 7) was evaluated by determining if the words were specific and varied and if the expressions and language used caught the reader’s interest. Each essay was scored by two independent raters who were trained in administration and scoring procedures of the WIAT-II to yield an overall score (a composite from each of the categories), and individual category scores. Raters were blind to the essay writer’s group membership and the time condition under which each essay was written. Inter-rater reliability for overall WIAT-II Essay Composition scores was good ( $r = .77$ ) and consistent with the overall inter-rater reliability estimates reported in the manual, but no values are reported in the manual by category. We computed the following estimates across the categories: mechanics  $r = .85$ ; organization  $r = .77$ , theme development  $r = .43$ , and vocabulary  $r = .43$ . Scoring the latter two categories involves markedly more subjectivity yielding lower inter-rater reliabilities, a common concern in analytic scoring of direct writing assessments (Johnson, Penny & Gordon, 2000; Shohamy, Gordon, & Kraemer, 1992). However, discrepancies between the raters were resolved through discussion until there was total agreement. Discussion is considered a viable resolution option when inter-rater agreements are low, especially when the stakes for the writing assessment are high (Johnson, Penny, Gordon, Shumate, & Fisher, 2005). A third rater (i.e., the *tertium quid model* of score resolution; Johnson et al., 2005) was also present to resolve discrepancies between scores when needed.

Essays were also scored across the same indices of lexical complexity Gregg et al. (2007) used in their study: (1) the number of different words (i.e., lexical diversity), (2) number of words with more than two syllables, and (3) average word length. Lexical diversity was also subjected to a more stringent analysis which accounts for text length using the Measure of Lexical Textual Diversity (MLTD; McNamara, Graesser, McArthur & Cai, 2014). Total word count and average word lengths were calculated with a computer program (Word Counter, 2016), and double checked for 100% accuracy by a research assistant. Syllable counts were computed by two independent raters who achieved 100% agreement.

### **Procedure**

Participants completed all of the tasks individually in one 90-minute session in a research lab on the university campus. Students completed two essay writing tasks by hand (Prompt A or B) counterbalanced across prompts and time. In between each of the essay writing tasks, students also completed the collection of cognitive and linguistic measures in a fixed order. For each writing task, students were informed of the time limit prior to beginning, and were provided with reminders about the time remaining at the five minute and one minute mark as they reached the end of their writing time. If students finished their essay before the allotted time, they were encouraged to keep writing until the time had expired. A research assistant was present in the room at all times to administer the measures, answer any questions the student had, or to provide participants with a short break if requested. Participants also completed a brief background questionnaire to provide confirmation on their diagnoses, their academic programs, and current access to academic accommodations.

## **RESULTS**

Preliminary analyses examined group differences on the cognitive, linguistic, and literacy measures and our main analyses examined students' essays across several dimensions when written under regular and extended time conditions. Specifically, this latter set of analyses addressed our research questions concerning group differences in writing quality across the various writing indices outlined previously and between time conditions, and whether evidence of a *differential* boost or an *interaction* between LD status and time would be found. Bonferroni correction was applied across all multiple comparisons to control for Type 1 error.

A series of one-way analysis of variance (ANOVA) were run to examine group differences on the cognitive, linguistic, and literacy measures. As shown on Table 1, the LD group performed significantly lower than the non-LD group on the measures of verbal WM  $F(1, 36) = 13.17, p = .001$  ( $\eta^2 = .27$ ) and RAN Letters,  $F(1, 36) = 7.82, p = .008$  ( $\eta^2 = .18$ ). There were no group differences on the expressive vocabulary measure. The next set of one-way ANOVAs examined group differences on the literacy measures. Table 1 summarizes the descriptive statistics across the reading (word reading and decoding fluency) and writing measures (spelling, handwriting fluency). As shown in Table 1, the LD group performed significantly lower on the TOWRE decoding fluency  $F(1, 36) = 28.58, p < .001$ , ( $\eta^2 = .44$ ), WRAT-3 Spelling  $F(1, 36) = 50.85, p < .001$  ( $\eta^2 = .59$ ) and handwriting fluency  $F(1, 36) = 13.24, p = .001$  ( $\eta^2 = .27$ ) measures.

**Table 1. Descriptive Statistics Across the Cognitive, Linguistic, and Literacy Measures**

Measure	LD (n = 19)		Non-LD (n = 19)		F	p
	M	SD	M	SD		
RAN – Digits	8.31	3.23	10.63	2.31	6.44	.016
RAN – Letters	7.21	2.57	9.73	2.97	7.82*	.008
Digit Span Forwards	10.57	2.11	11.00	2.35	.366	.566
Digit Span Backwards	6.42	1.30	8.47	2.09	13.17*	.001
Vocabulary	12.15	1.89	11.68	1.60	.694	.410
TOWRE – Word Reading	91.57	11.23	99.89	10.94	5.33	.027
TOWRE – Decoding Fluency	88.78	8.81	105.52	10.42	28.58*	<.001
WRAT-3 – Spelling	94.10	9.80	111.42	3.97	50.85*	<.001
Handwriting Fluency	86.63	18.81	108.52	18.26	13.24*	.001

Note. Scores are based on standard scores and scaled scores except for Digit Span Forwards, Backwards and Handwriting Fluency which are raw score. RAN = Rapid Automatized Naming, TOWRE = *Test of Word Reading Efficiency*; WRAT-3 = Wide Range Achievement Test-3rd Edition. \* $p < .008$ .

The results of a 2 (group: LD, non-LD) X 2 (time: 15, 30 minutes) repeated measures ANOVA for the overall essay scores (a composite of mechanics, organization, theme development and vocabulary) indicated a significant main effect for group  $F(1, 36) = 13.01, p = .001$  ( $\eta^2_{\text{partial}} = .27$ ), with the LD group achieving lower overall essay writing scores than the non-LD group irrespective of the time provided. Follow-up one-way ANOVAs examined the group differences in essay performance across the two time conditions (regular and extended time) for each of the essay components (mechanics, organization, theme development, vocabulary). As shown in Table 2, this analysis under the regular time conditions indicated that the LD group scored significantly lower in the area of mechanics than the non-LD group producing more spelling and punctuation errors in their writing  $F(1, 36) = 12.43, p = .001$  ( $\eta^2 = .26$ ). There were no other significant differences between groups across the other components of writing under regular time conditions.

We conducted additional one-way ANOVAs to examine group differences across the components of writing under the extended time condition. As shown in Table 2, the results indicated that the LD group scored significantly lower in the area of mechanics than the non-LD group,  $F(1, 36) = 9.86, p = .003$  ( $\eta^2 = .22$ ), similar to the results under the regular time condition. However, the results also indicated that the students with LD wrote significantly less organized essays under the extended time condition,  $F(1, 36) = 10.42, p = .003$  ( $\eta^2 = .22$ ), compared to the non-LD group suggesting that extended time may have actually disadvantaged the quality of these students’ essays. No other significant group differences on these components of writing were found under the extended time condition.

The length of texts students produced under regular and extended time conditions was examined. A 2 (group) X 2 (time) repeated measures ANOVA examining



word count as a measure of text length revealed a main effect of time,  $F(1, 36) = 18.13, p = .001$  ( $\eta^2_{\text{partial}} = .34$ ), but no interaction although a trend to significance was evident ( $p = 0.67$ ). Both groups produced more text under the extended time conditions. However, the average number of additional words written by the LD group with extended time was nearly twice as much as the additional words written by the non-LD group with extended time, and more similar to the text production of the non-LD group under regular time. To follow-up, two one-way ANOVAs examined group differences across text length for each time condition. As shown in Table 3, these analyses indicated that the LD group wrote significantly fewer words than the non-LD group  $F(1, 36) = 8.90, p = .005$  ( $\eta^2 = .20$ ) when provided with regular time; however, there was no significant difference between the groups for the 30 minute essay ( $p = .88$ ; see Table 3). Paired samples  $t$ -tests also indicated that the difference in text length between regular and extended time was significant for the LD group  $t(18) = -3.59, p = .002$ , but not for the non-LD group. We also calculated Cohen's  $d$  to determine effect size and found an effect size of 0.82, suggesting a larger effect of extended time for students with LD's performance when it came to text length. For the non-LD group, we found an effect size of 0.53, suggesting only a medium effect of extended time on performance. Taken together, these findings suggest that when provided with additional time, students with LD are able to write as much as their non-LD peers, removing the differences in text length between groups indicative of a *differential boost* in performance.

As per the approach used by Gregg et al. (2007), we also computed three indices of lexical complexity: (1) number of different words (i.e., lexical diversity), (2) number of words with more than two syllables, and (3) average word length to see whether there would be group differences across these indices, and to see whether students would produce more sophisticated word choices in writing when provided with more time. The first index of lexical complexity was reflected in the number of different words produced by students in their essays under regular and extended time conditions. A 2 (group) X 2 (time) repeated measures ANOVA showed a main effect of time,  $F(1, 36) = 13.94, p = .001$  ( $\eta^2_{\text{partial}} = .28$ ), but no interaction although a trend to significance was evident ( $p = 0.69$ ). As shown in Table 3, follow-up one-way ANOVAs indicated that the LD group wrote significantly fewer unique words than the non-LD group  $F(1, 36) = 9.21, p = .004$  ( $\eta^2 = .20$ ) when provided with regular time. However, there was no significant difference between the groups for the number of different words written on the 30 minute essay ( $p = .89$ ). Paired samples  $t$ -tests indicated a significant difference for the LD group only,  $t(18) = -3.27, p = .004$  (Cohen's  $d = .75$ ) who made significantly more varied word choices in their writing under the extended rather than under the regular time condition, suggestive of a *differential boost* in performance. However, when lexical diversity was scored according to the more stringent MTL approach (which controls for text length), repeated measures results indicated no main effects for group or time, and no interactions. Within-group  $t$ -test comparisons were also not significant. The fact that the LD group produced less text under regular time conditions, reflecting a core feature of their processing and skills deficits, raises questions about the validity of the MTL approach when comparing lexical diversity in students with and without LD. This point will be discussed more in the subsequent section.

**Table 2. Group Differences on Essay Components (WIAT-II Scoring) – 15 and 30 Minutes**

Measure	15 Minute Essays						30 Minute Essays						
	LD			Non-LD			LD			Non-LD			
	M	SD	F	M	SD	p	M	SD	F	M	SD	F	p
Overall Score	21.05	4.84	9.91*	25.89	4.62	.003	20.89	3.64	10.28*	25.84	5.64	10.28*	.003
Mechanics	3.15	2.40	12.43*	5.57	1.77	.001	2.42	2.31	9.86*	4.84	2.43	9.86*	.003
Organization	10.10	2.25	4.09	11.7	2.83	.051	9.94	1.74	10.42*	12.21	2.50	10.42*	.003
Theme Development	4.94	1.31	.528	5.26	1.36	.472	5.73	1.40	1.61	5.21	1.13	1.61	.213
Vocabulary	2.84	1.21	.944	3.26	1.44	.338	2.78	1.18	2.48	3.57	1.83	2.48	.124

Note. Scores based on WIAT-II Essay Composition Raw Scores. WIAT-II = Wechsler Individual Achievement Test Second Edition. \* $p < .01$ .

**Table 3. Group Differences on Additional Essay Components – 15 and 30 Minutes**

Measure	15 Minute Essays						30 Minute Essays						
	LD			Non-LD			LD			Non-LD			
	M	SD	F	M	SD	p	M	SD	F	M	SD	F	p
Text Length	159.11	53.78	8.90*	207.79	46.58	.005	239.21	97.81	.023	243.89	93.53	.023	.881
Lexical diversity	99.95	22.49	9.21*	121.26	20.78	.004	134.79	44.99	.021	132.84	36.81	.021	.885
Words > 2 syllables	20.37	13.00	5.18	28.42	8.29	.029	33.00	19.45	.355	36.53	16.95	.355	.555
Word length	4.70	0.35	3.98	4.90	0.27	.054	4.76	0.28	3.98	4.94	0.28	3.98	.054

Note. Text length = word count, lexical diversity = number of different words.

The second index of lexical complexity was derived from the number of words produced in students' essays that contained more than two syllables. A 2 (group) X 2 (time) repeated measures ANOVA indicated a main effect for time,  $F(1, 36) = 9.43, p = .004$  ( $\eta^2_{\text{partial}} = .21$ ), but no interaction ( $p = 0.51$ ). Both groups produced more words with more than 2 syllables under the extended time conditions. The final index of lexical complexity was computed based on the average length of the words students wrote. A 2 (group) X 2 (time) repeated measures ANOVA examining average word length revealed a main effect of group,  $F(1, 36) = 7.46, p = .01$  ( $\eta^2_{\text{partial}} = .17$ ) with students with LD writing essays with shorter words than their non-LD peers irrespective of time.

## DISCUSSION

This study is the first to employ a repeated measures design to examine the efficacy of extended time on the writing performance of university students with LD. This design afforded a direct evaluation of whether extended time significantly improves only the performance of students with LD (a test of the interaction hypothesis), or improves the quality of essays produced by all students, but increases the scores obtained the LD group more significantly (a test of the differential boost theory). Overall, our results indicated no evidence to support the more conservative interaction hypothesis, although an interaction effect on the measure of text length (based on word count) did trend in this direction ( $p = .067$ ). We did find some evidence to support the theory of a differential boost on text length and with one of the measures of lexical diversity. Based on the analysis of differences in word count with extended time, the LD group wrote nearly twice as many more words as the non-LD group with extended time. These results parallel Lewandowski et al's. (2008) findings that extended time provides students the opportunity to attempt as many questions as students without LD in the allotted time. Both groups in our study produced more text under extended time, but the LD group produced significantly more text suggesting a *differential boost* in performance. Similarly, while both groups produced more lexically diverse text (as a measure of lexical complexity) when provided more time to write, within-group comparisons indicated that the LD group's text contained a significantly greater number of *different* words written with extended time. McCarthy & Jarvis (2010) have reported another more conservative approach to measuring lexical diversity (measure of textual lexical diversity; MTLT). This scoring method controls for text length as a possible confound to writers using more diverse word choices in their written text. The premise is based on the finding that in typical writers as texts lengthen, word choices become less diverse and more homogeneous to adhere to a coherent theme or topic (McCarthy & Jarvis, 2010). A central individual difference distinguishing writers with and without LD is the length of the texts they produce, as we also replicated in the present study. In comparing LD and non-LD writers, controlling the lengths of the texts they produce actually *removes* one of the core processing differences between the groups; that is, the differences in automaticity for lower-order writing skills. None of the studies published thus far using the MTLT approach have been conducted comparing LD and non-LD writers. Moreover, in keeping with the rationale for using such an approach, we actually have further evidence to augment the finding that the LD writers wrote texts that were less-organized with

more time. Their tendency to use more lexically diverse words with extended time compared to regular time suggests that the more they wrote, the more diverse their word choices, and hence they were less likely to “stick with a topic or theme”. More research on the validity of the MTL approach in studies comparing LD and non-LD writers is certainly warranted.

With more time, students with LD are writing more, and using a greater variety of word choices in their writing with extended time. These results are consistent with the studies of Alster (1997) and Lesaux et al., (2006) where students with LD experienced a significantly greater magnitude in score changes when provided with extended time, which can be taken as some evidence for *differential boost*. This differential boost in performance did not extend to the other components of lexical complexity where the LD group produced less sophisticated words in text (based on the use of words with more than two syllables, or producing on average words of greater length), irrespective of time condition. Gregg et al. (2007) also reported that the LD group in their study were distinguished from the non-LD group on the same measures of text length (i.e. word count) and on two of the three measures of lexical complexity, based on one essay task administered under standard time conditions across groups.

We also found no evidence for the efficacy of extended time on overall writing quality when evaluated across lower order and higher order aspects. Regardless of time, the LD group achieved significantly lower scores overall and across the lower-order (mechanics) and higher-order (theme development, organization, vocabulary) aspects of essay quality. In fact, as already noted, there was evidence to suggest that being provided with more time may actually have disadvantaged the LD group in relation to the organization of their essays. The LD group wrote more and used more lexically diverse words in their writing with extended time, but they also had more opportunity to write more disorganized texts—failing to apply strategies (if such strategies are even within their repertoire) integral to the self-regulation of the writing process such as planning, monitoring, and revising (Harris & Graham, 1996). The LD group performed significantly below the non-LD group on the working memory (WM) measure. WM is a critical component of contemporary writing models where multiple demands across higher and lower level aspects of the writing process must be simultaneously coordinated (Berninger & Amtmann, 2003; Olive & Kellogg 2002). The greater quantity of text students produced with more time likely put a greater toll on working memory resources impacting the quality of text organization.

We found low inter-rater reliability estimates, however, on two of the categories (theme development and vocabulary) included in the WIAT-II. While we resolved this discrepancy to total agreement through discussion, the fact that even with rubrics provided for analytic scoring of essays, choice of score resolution (Johnson, Penny, & Gordon, 2000) in reliably evaluating writing quality—for future research, and in diagnostic assessment of students with LD in higher education.

In addition to significant differences in WM between the LD and non-LD groups, the LD group also achieved significantly lower scores than the non-LD group, and scores clinically below age-expectations, on the naming speed (RAN) and pseudoword decoding measures. Thus, persistent processing deficits remain with these critical predictors of literacy performance, and reflect profiles consistent with previ-

ous research (e.g., Parilla, Georgiou, & Corkett, 2007). The LD group also underperformed compared to the non-LD group on the spelling and handwriting fluency measures also consistent with previous research with this population (e.g., Connelly et al., 2006; Harrison & Beres, 2007; Harrison, 2009.). Students' expressive vocabulary and word reading fluency, however, did not significantly differ from the non-LD group. This pattern of performance suggests that our sample of students with LD may be considered more similar to the student with "high functioning" dyslexia (Deacon, Parilla, & Kirby, 2006). Most of the students with LD in our sample were well into their 4-year undergraduate programs, and thus had accumulated some academic success to persist in their respective programs. They were also well-connected with the Disability Resource Centre on campus that runs programs to promote learning strategies and self-advocacy. As Getzel (2008) noted, settings that provide students with a suite of services that promote their self-determination, strategy development, and autonomy within the context of accommodating their learning needs in higher education lead to a greater retention of undergraduates with LD.

Some potential limitations with our study are important to note. The assessment was conducted within one 90-minute session, and while possible fatigue could have resulted, this does not seem likely as students are often required to write exams for multiple hours. Furthermore, students were able to take a break if requested. Students were not provided with an exam setting in which they were to write their essays, and this could have impacted the results. Not only did students not experience the potential heightened anxiety often associated with exam writing, but they also did not respond to prompts for which they were required to study. Future research might simulate these contextual experiences to see how performance is impacted when students are provided with additional time. Furthermore, many students receive additional accommodations such as the use of assistive technology or a separate setting for tests. Therefore, students with LD might require the provision of all of their accommodations to truly perform at their *maximum potential* or demonstrate a *differential boost* in their scores. Other accommodations may impact the performance of students within a testing situation in conjunction with extended time. Future research might examine multiple accommodations and their relations to student performance.

Additional research across higher education settings with a larger sample of students than the current study, and comparing a variety of accommodations (e.g., separate setting, the use of technology to write) is needed to provide further insight into the impact of certain accommodations on academic performance in students with LD. Ultimately, we found limited evidence that students with LD would demonstrate improved writing performance with extended time when examined against two prominent theories: differential boost or interaction hypothesis. Academic accommodations may "level the playing field" in relation to task demands for students with LD, but academic skills difficulties and processing deficits persist irrespective of the accommodation of extra time, as noted by others (Gregg & Coleman, 2009). Interventions aimed at improving persistent skills difficulties, such as in writing, continue to be necessary for students with LD in higher education.

## REFERENCES

- Berninger, V. W. (1999). Coordinating transcription and text generation in working memory during composing: automatic and constructive processes. *Learning Disabilities Quarterly*, 22, 99-112. Retrieved from <http://www.jstor.org/stable/1511269>
- Berninger, V. W., & Amtmann, D. (2003). Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of Learning Disabilities* (pp. 345-363). NY: Guilford Press.
- Berninger, V. W., Mizokawa, D., & Bragg, R. (1991). Theory based diagnosis and remediation of writing disabilities. *Journal of School Psychology*, 29, 57-79.
- Connelly, V., Campbell, S., MacLean, M., & Barnes, J. (2006). Contribution of lower order skills to the written composition of college students with and without dyslexia. *Developmental Neuropsychology*, 29, 175-196.
- Connelly, V., Dockrell, J., & Barnett, J. (2005). The slow handwriting of undergraduate student constrains overall performance in exam essays. *Educational Psychology*, 25, 99-107.
- Deacon, S., Parilla, R., & Kirby, J. R. (2006). Processing of derived forms in high-functioning dyslexics. *Annals of Dyslexia*, 56, 103-128.
- Fuchs, L. S. & Fuchs, D. (2001). Helping teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research & Practice*, 16, 174-181.
- Getzel, E.A. (2008). Addressing the persistence and retention of students with disabilities in higher education: Incorporating key strategies and supports on campus. *Exceptionality*, 16, 207-219.
- Gregg, N., & Nelson, J. M. (2012). Meta-analysis on the effectiveness of extra time as a test accommodation for transitioning adolescents with learning disabilities: More questions than answers. *Journal of Learning Disabilities*, 45, 128-138.
- Graham, S., & Weintraub, N. (1996). A review of handwriting research: Progress and prospects from 1980 to 1984. *Educational Psychology Review*, 8, 7-87.
- Gregg, N., Coleman, C., Davis, M., & Chalk, J. C. (2007). Timed essay writing: Implications for high-stakes tests. *Journal of Learning Disabilities*, 40, 306-318.
- Gregg, N., Hoy, C., McAlexander, P, & Hayes, C. (1991). Written sentence production error patterns of college writers with learning disabilities. *Reading and Writing: An Interdisciplinary Journal*, 3, 169-185. Alexander & Hayes (1991)
- Harrison, G. L. (2009). The component reading and writing skills of at-risk undergraduates with learning difficulties. *Learning Disabilities: A Contemporary Journal*, 7(2), 59-72.
- Harrison, G. L., & Beres, D. (2007). The writing strategies of post-secondary students with learning disabilities. *Exceptionality Education Canada*, 17(2), 221-242.
- Hatcher, J., Snowling, M. J., & Griffiths, Y. M. (2002). Cognitive assessment of dyslexic students in higher education. *British Journal of Educational Psychology*, 72, 119-133.
- Johnson, R. L., Penny, J., & Gordon, B. (2000). The relation between score resolution methods and interrater reliability. An empirical study of an analytic scoring rubric. *Applied Measurement in Education*, 13, 121-138.
- Johnson, R. L., Penny, J., Gordon, B., Shumate, S. R., & Fisher, S. P. (2005). Resolving score differences in the rating of writing samples: Does discussion improve the accuracy of scores? *Language Assessment Quarterly*, 2, 117-146.
- Lesaux, N. K., Pearson, M. R. & Siegel, L. S. (2006). The effects of timed and untimed testing conditions on the reading comprehension performance of adults with reading disabilities. *Reading and Writing*, 19, 21-48.
- Lewandowski, L. J., Lovett, B. J., & Rogers, C. (2008). Extended time as a testing accommodation for students with reading disabilities: Does a rising tide lift all ships? *Journal of Psychoeducational Assessment*, 26, 315-324.

- Li, H., & Hamel, C. M. (2003). Writing issues in college students with learning disabilities: A synthesis of the literature from 1990-2000. *Learning Disability Quarterly*, 26, 29-46. Retrieved from <http://www.cldinternational.org/Publications/LDQ.asp>
- Lindstrom, J. H. (2007). Determining appropriate accommodations for postsecondary students with reading and written expression disorders. *Learning Disabilities Research & Practice*, 22, 229-236.
- Lovett, B. J. (2010). Extended time testing accommodations for students with disabilities: Answers to five fundamental questions. *Review of Educational Research*, 80, 611-638.
- Lovett, B. J. (2011). Extended Time Testing Accommodations: What Does the Research Say?. *Communique: The newspaper of the National Association of School Psychologists*, 39, 14-15. Retrieved from [http://sss.eportalnow.net/uploads/1/1/7/2/11720827/extended-time\\_accommodations.pdf](http://sss.eportalnow.net/uploads/1/1/7/2/11720827/extended-time_accommodations.pdf)
- McCarthy, P. M., & Jarvis, S. (2012). MTLTD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42, 381-392.
- McGonnell, M., Parilla, R., & Deacon, S. H. (2007). The recruitment and description of university students who self-report difficulty acquiring early reading skills. *Exceptionality Education Canada*, 17, 155-174.
- McNamara, D. D., Graesser, A. C., McCarthy, M., & Cai, Z. (2014). Automated evaluation of text and discourse with Coh-Metrix. New York: Cambridge University Press.
- Ofiesh, N. S. (2000). Using processing speed tests to predict the benefit of extended test time for university students with learning disabilities. *Journal of Postsecondary Education and Disability*, 14, 39-56. Retrieved from <http://www.eric.ed.gov/PDFS/EJ649045.pdf>
- Ofiesh, N. S., Hughes, C. A. (2002). How much time?: A review of the literature on extended test time for postsecondary students with learning disabilities. *Journal of Postsecondary Education and Disability*, 16, 2-16. Retrieved from [http://www.kennesaw.edu/stu\\_dev/dsss/howmuchtime.pdf](http://www.kennesaw.edu/stu_dev/dsss/howmuchtime.pdf)
- Ofiesh, N. S., Hughes, C., & Scott, S.S. (2004). Extended test time and postsecondary students with learning disabilities: A model for decision making. *Learning Disabilities Research & Practice*, 19, 57-70. Retrieved from <http://www.wiley.com/bw/journal.asp?ref=0938-8982>
- Olinghouse, N. G. & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing: An Interdisciplinary Journal*, 26, 45-65.
- Olive, T., & Kellogg, R. T. (2002). Concurrent activation of high- and low-level production processes in written composition. *Memory and Cognition*, 30, 594-600.
- Parilla, Georgiou, & Corkett (2007) University students with a significant history of reading difficulties: What is and is not compensated? *Exceptionality Education International*, 17, 195-220.
- Shohamy, E., Gordon, C. M., & Kraemer, R. (1992). The effect of raters background and training on the reliability of direct writing tests. *The Modern Journal*, 76, 27-33.
- Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75, 457-490.
- Stretch, L. S. & Osborne, J. W. (2005). Extended time test accommodation: Directions for future research and practice. *Practical Assessment, Research & Evaluation*, 10, 1-8.
- The Psychological Corporation (2002). *Wechsler Individual Achievement Test-2<sup>nd</sup> Edition*. Toronto, ON: Harcourt Canada.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: ProEd.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *The Comprehensive Test of Phonological Processing*. Austin, TX: ProEd.

- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-3<sup>rd</sup> Edition*. San Antonio, TX: The Psychological Corporation.
- Weis, R., Dean, E. L., & Osborne, K. J. (2014). Accommodation decision making for postsecondary students with learning disabilities: Individually tailored or one size fits all? *Journal of Learning Disabilities, 49*, 484-498.
- Wilkinson, G. S. (1993). *The Wide Range Achievement Test-3<sup>rd</sup> Edition*. Wilmington, DE: Jastak Associates. *Journal of Learning Disabilities, 34*, 394-400.
- Word Counter (2016). Word Count Tool. Retrieved from <https://wordcounttools.com>.
- Zuriff, G. E. (2000). Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Applied Measurement in Education, 13*, 99-117.