Effects of Extended Time Allotments on Reading Comprehension Performance of College Students With and Without Learning Disabilities

Lawrence Lewandowski¹, Justin Cohen¹, and Benjamin J. Lovett²

Abstract

Students with disabilities often receive test accommodations in schools and on high-stakes tests. Students with learning disabilities (LD) represent the largest disability group in schools, and extended time is the most common test accommodation requested by such students. This pairing persists despite controversy over the validity of extended time as a test accommodation. The current study examined the effects of 50% and 100% time extensions on the reading comprehension performance of college students with and without LD. Results indicated that typical students actually benefited more than the LD group when given extra time, indicating that extended time is not a test accommodation that is specific to those with a disability. Moreover, when only students with LD were given extended time, especially double time, they outperformed nondisabled peers. We discuss implications of these findings for future research as well as accommodation decisions in educational settings.

Keywords
test accommodations, extended time, learning disabilities

Students with learning disabilities (LD) make up approximately 50% of the entire special education population in the United States (U.S. Department of Education, National Center for Educational Statistics, 2010). As large numbers of students with disabilities enter postsecondary education, many of these students apply for test accommodations in these settings as well as on high-stakes examinations such as the Scholastic Aptitude Test (SAT), Graduate Record Examination (GRE), and the Law School Admission Test (LSAT; Cahalan-Laitusis, King, Cline, & Bridge-man, 2006; Ofiesh, 2006). Test accommodations are changes in the standard administration of a

¹Syracuse University, Syracuse, NY, USA
²Elmira College, Elmira, NY, USA

Corresponding Author:
Lawrence Lewandowski, Department of Psychology, 472 Huntington Hall, Syracuse University, Syracuse, NY 13244, USA.
Email: ljlewand@syr.edu
test including testing procedures or formats that enable students with disabilities to participate in assessment programs on an equal basis with their nondisabled peers (Sireci, Scarpati, & Li, 2005). Specifically, extended time is one of the most frequently requested and granted test accommodation, and it is available on most state-wide assessments as well as on the previously noted national high-stakes exams (Bolt & Thurlow, 2004; Stretch & Osborne, 2005). Despite the wide-scale use of test accommodations such as extended time, there have been controversies as to its validity as well as who should qualify for these provisions (Lerner, 2004; Lovett, 2010).

In theory, extra time to complete examinations is a logical accommodation to provide to students with LD, due to these students’ deficits in processing speed (Bryant, McIntyre, Murray, & Blackwell, 1983; Weiler et al., 2000), and difficulty completing timed tests (Cahalan-Laitusis et al., 2006). However, the fairness of extended time as an accommodation has been questioned. Researchers have found in several studies that students without disabilities benefit from extended time (see Sireci et al., 2005, for review), and in some cases benefit even more than students with LD (Lewandowski, Lovett, & Rogers, 2008). According to Phillips (1994), an accommodation should mitigate the impairment of the disability but not improve the performance of students without disabilities. These findings raise concerns about the validity of extended time as an accommodation, and have spawned research on what has been termed the “interaction hypothesis.” Applied to extended time, the interaction hypothesis predicts that LD and nondisabled groups would differ more at standard time than at extended time because the LD students would “catch up” (Sireci et al., 2005).

Phillips (1994) has raised additional concerns about the validity of test accommodations. For instance, she has argued that a test accommodation should only be given to a person with an accurate disability diagnosis. There has been much debate about how LD should be diagnosed (see chapters in Flanagan & Alfonso, 2011), and researchers have found that diagnoses are sometimes conferred without individuals meeting full criteria for the diagnosis, especially in college populations (Harrison, Nichols, & Larochette, 2008; Sparks & Lovett, 2009). In particular, some scholars have questioned whether or not diagnosticians have applied the “impairment criterion” to their diagnoses, such that the individual with a disability is substantially limited in academic skills relative to the average person (e.g., Lovett & Lewandowski, 2006). If a nonimpaired student receives a test accommodation, then he or she may actually be given an advantage over peers taking the exam under standard conditions.

Another concern of Phillips (1994) is that the test accommodation should not alter the test or change the meaning of the test score so as to make it less indicative of the underlying construct being measured. For instance, having a test read to a student in a separate room while using double time and a scribe may alter the test so dramatically that it is not measuring the same construct(s) as the standard version of the test. In fact, two studies already have shown that extended time alone has been associated with lowering the predictive validity of tests (Cahalan, Mandinach, & Camara, 2002; Thornton, Reese, Pasheley, & Dalessandro, 2001). It seems plausible that some students receive extended time even though (a) they do not meet all diagnostic criteria, (b) they are not significantly impaired relative to peers, (c) the altered test procedure may overpredict their performance, and (d) other students would benefit from the accommodation. In such a case, the use of this accommodation would seem to increase rather than decrease the score error, and this is certainly not the intent of legal mandates that provide test accommodations for students with clear disabilities and attendant impairment. However, even when students have accurate diagnoses, demonstrate impairment, and qualify for test accommodations, there is no empirical guidance as to what accommodations in what amount would be most appropriate.

Phillips (1994) also noted the need for reliable decision procedures about accommodations. However, decisions about extended time involve apparently arbitrary amounts of additional time being given to examinees. Although varying levels of extended time, including unlimited time,
are requested by students with disabilities, practitioners, and tests administrators have little empirical basis on which to make extended time allotment decisions. Extended time is provided in various amounts, most commonly as 50% or 100% extra time. These time amounts are provided as standard options on applications for tests such as the SAT, ACT, and GRE. Extensions of 50% to 100% time appear to be the default accommodations for these tests, and Ofiesh and Hughes (2002) suggested that extensions in this range be provided as a “general rule.” However, there is currently no standardized method for determining the amount of extended time to grant students with disabilities, and these decisions are made on a case-by-case basis relying largely on clinical judgment. There is, then, no scientific basis for providing LD students with 50% or 100% more time than peers, as opposed to 25% or 75%, for example.

How much time does a student with LD need to finish an exam? Or, more specifically, how much time is needed for that student to access as much of the exam as nondisabled peers? The research literature is rather silent on this issue. Researchers at the College Board have examined this question in the case of the SAT. Cahalan-Laitusis et al. (2006) investigated the use of extended time by high school juniors and seniors with LD and/or Attention Deficit Hyperactivity Disorder (ADHD) on sections of the SAT. Depending on the section of the test, they found that students with disabilities used an average of only 8% to 14% more time than peers to complete the subtests. However, these results may not generalize beyond the SAT, a lengthy test during which fatigue may keep students from fully utilizing their accommodations. Clearly there is a need for research that begins to formulate an empirical basis for extended time allotment decisions.

The current study examined the reading comprehension skills of college students with and without LD under three time conditions: standard time, time and a half (50% extended time), and double time (100% extended time). The primary aim of this study was to test the interaction hypothesis. This posits that non-LD students should outperform LD students at standard time, but the groups will perform similarly when they both receive extended time (i.e., time and one-half). In addition, we compared the non-LD group performance at standard time to the LD group performance at time and one-half and double time. In addition to number of comprehension items answered correctly, we also examined number of items attempted to determine the amount of work output generated by each group at each time interval.

**Method**

**Participants**

A total of 107 students enrolled at two private four-year colleges in Upstate New York were recruited from undergraduate classes and via flyers posted on campus. Students who reported LD status were considered eligible if they had access to test accommodations through the campus office of disability services (ODS). ODS policy requires a recent professional report (within 3 years of accommodations request) that includes a *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV)* diagnosis of LD. Twenty-six LD students met these criteria, and 13 of the 26 students reported a comorbid condition (i.e., ADHD, anxiety, depression). An attempt was made to select a “real-world” college LD sample in which comorbidity is common and everyone receives extended time. By contrast, students without disabilities had to be a native English language speaker/reader, free of any diagnosed learning or psychiatric disorder, and not eligible to receive test accommodations. A total of 81 students were screened and 50 students met these criteria.

The average age of LD participants was 19.7 (SD = 1.78) with a range of 18 to 26 years; and the average age of students without disabilities was 19.1 (SD = 1.02), ranging from 18 to 23
years. There was a slightly higher percentage of juniors and seniors in the LD group (27% vs. 16%). The average reported grade point average (GPA) for the LD students was 2.91 ($SD = 0.64$), while the average GPA of students without disabilities was 3.28 ($SD = 0.57$). Fifty percent of the LD group and 57.7% of the non-LD group were female. Sixty-seven percent of the non-disabled group and 84.6% of the LD group were Caucasian. LD participants reported primary disabilities in reading (24) and writing (2), with 5 students indicating a concomitant learning disability in math. Students with LD in writing, known to have problems with literacy, had similar profiles to the LD reading sample and were included in analysis. Students with dual disabilities in reading and math also showed no differences in test profiles and were included in analysis. All LD students reported receiving extended time as an accommodation on any and all tests regardless of content.

**Measures**

The main measure used in this study is a modified version of the reading comprehension subtest of the Nelson–Denny Reading Test (NDRT; Brown, Fishco, & Hanna, 1993). The NDRT has two parallel forms, G and H. Each reading comprehension form consists of seven passages, accompanied by 38 multiple-choice questions. These questions are both factual and inferential in nature. According to the NDRT technical manual, the Nelson–Denny reading comprehension subtest has alternate forms reliability of 0.81 and an internal consistency of 0.88 for forms G and H. This test has been used in dozens of studies with poor and LD readers, including many of the reading studies on extended time.

For this study, forms G and H of the NDRT reading comprehension subtest were combined in order to make a 76-item test. Students completed form H, followed by form G. We purposely increased the amount of test content to avoid potential ceiling effects as we extended time. Pilot testing found that the typical administration conditions of the NDRT would need to be modified; this is not surprising, given that in diagnostic use, the NDRT is used for students as young as ninth graders. In the present study, each participant’s modified NDRT was scored to calculate his or her number of correct items, the number of items attempted, and percentage correct, at each of three time points (see below).

A short demographics form was included at the end of the session. Each participant was asked to provide information on age, gender, ethnicity, school year, GPA, disability status, and accommodations received on tests.

**Procedure**

Measures were administered to groups of participants consisting of between one and eight students. Two students with LD and one without LD were tested individually due to scheduling conflicts. Students read an informed consent that explained the purpose of the study and the tasks involved, during which any questions were addressed. Following consent, all participants were provided with a packet of the materials and a red pencil. All students were informed that they would be receiving extra time to complete the exam and to work to the best of their ability as though the current test was a high-stakes assessment such as the SAT. Participants started with red-colored pencils and were given exactly 15 min to read and answer multiple-choice questions. After 15 min, the examiner required all students to stop working and exchanged each participant’s red pencil for a blue pencil. The examiner then allowed the students to resume work for an additional 7 min and 30 s (time and one half). After this, the examiner again stopped all participants, collected the blue pencils, and distributed green pencils. Students were then prompted to begin the final testing phase for an additional 7 min and 30 s. After a total of
30 min (double time) to work on the exam, all participants were instructed to stop and asked to complete the brief demographics questionnaire.

During the testing students were free to use any testing strategy or approach, including revisiting and revising items already completed. Procedural integrity was evaluated for 38% of data collection sessions. Three doctoral-level graduate students observed the examiner on a number of occasions to ensure that a standardized protocol was followed with regards to instructions and timing. According to checklists completed by the three observers, 99% of the observed sessions involved all standardized procedures being followed.

Results

Reliability

In order to ensure that the modified measure used in the current study had similar reliability to that of the Nelson−Denny reading comprehension subtest, a split-half reliability coefficient (odd items vs. even items) was calculated. Twenty-five protocols from the nondisabled group were randomly selected to be used as data in this analysis. The correlation between odd items and even items was \( r = 0.82 \), similar to the reliability data reported in the NDRT Technical manual (internal consistency reliability = 0.88). This suggests that the modified measure used in this study has adequate reliability.

Descriptive Statistics

Groups were examined to determine comparability on variables unrelated to disability status. \( \chi^2 \) tests were employed to compare group distributions on ethnicity and sex. The results of \( \chi^2 \) analyses indicated that there were no significant differences between the groups on ethnicity, \( \chi^2(7, 75) = 10.096, p = .121 \) or sex, \( \chi^2(1, 76) = 0.406, p = .524 \). Independent groups \( t \) tests were employed to check for differences on the continuous variables of age and GPA. There were no group differences for age, \( t(73) = 1.932, p = .057 \). The LD group had a lower average GPA than the nondisabled group, \( t(73) = -2.469, p < .05, d = 0.6 \), but this difference was not surprising given that LD is defined in part by difficulties in academic achievement.

Interaction Hypothesis Analysis

Table 1 presents the mean scores for the LD and nondisabled student groups on each of the dependent measures in the study at each time condition. The interaction hypothesis was tested via a \( 2 \times 3 \) (Group [LD, non-LD] × Time Condition [Standard, Time and One-Half, Double]) mixed model analysis of variance (ANOVA). Three ANOVAs were conducted for the dependent measures items correct, items attempted, and percentage correct/accuracy (see Table 2). According to the interaction hypothesis, the students with LD should perform significantly worse than peers at standard time and comparably to peers at time and one-half, resulting in a significant Group × Time interaction.

Items correct. The 2 Group × 3 Time Condition ANOVA for items correct yielded a significant interaction between time and group, \( F(1, 148) = 9.7, p < .001, \eta_p^2 = 0.116 \). However, this interaction was not in the predicted direction of the interaction hypothesis. The existing performance gap between the groups at standard time became wider over time. A main effect for group demonstrated that students without disabilities outperformed students with LD overall, \( F(1, 74) = 20, p < .01, \eta_p^2 = 0.213 \). A main effect for time condition showed that overall, all participants improved given extended time, \( F(1, 148) = 1.192, p < .001, \eta_p^2 = 0.94 \).
Planned comparisons were conducted between the performance of students without disabilities at 15 min (standard time) and the performances of students with LD at 15, 22.5, and 30 min. LD students answered significantly fewer items correct than their nondisabled peers at 15 min, \(t(74) = -4.413, p < .01, d = -1.1\). However, when the LD group received time and one-half, they outperformed their standard-time peers (LD \(M = 31.04\), non-LD \(M = 26.38\)), \(t(74) = 2.6, p = .011, d = 0.61\). This difference increased further when only the LD group was given double time (LD \(M = 42.08\), non-LD \(M = 26.38\)), \(t(74) = 7.4, p < .001, d = 1.65\). These comparisons are presented in Table 3.

**Items attempted.** The 2 Group × 3 Time Condition ANOVA for items attempted yielded a significant interaction between time and group, \(F(1, 148) = 8.88, p < .001, \eta^2_p = 0.107\). However, this interaction was again not in the predicted direction of the interaction hypothesis. The existing performance gap between the groups at standard time became increasingly wider with more extended time. A main effect for Group demonstrated that students without disabilities attempted more items than students with LD overall, \(F(1, 74) = 18.9, p < .01, \eta^2_p = 0.204\). A main effect for

### Table 1. Descriptive Statistics.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Nondisabled ((n = 50))</th>
<th>LD ((n = 26))</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard time (15 min)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Items correct</td>
<td>26.38 (6.99)</td>
<td>19.35 (5.73)</td>
<td>1.1</td>
</tr>
<tr>
<td>Items attempted</td>
<td>29.94 (7.06)</td>
<td>23.15 (6.09)</td>
<td>1.03</td>
</tr>
<tr>
<td>% Accuracy (correct/attempted)</td>
<td>88.1 (9.1)</td>
<td>83.5 (11)</td>
<td>0.44</td>
</tr>
<tr>
<td><strong>Time and a half (22.5 min)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Items correct</td>
<td>40.42 (9.76)</td>
<td>31.04 (8.23)</td>
<td>1.03</td>
</tr>
<tr>
<td>Items attempted</td>
<td>46.26 (9.76)</td>
<td>36.88 (9.30)</td>
<td>0.98</td>
</tr>
<tr>
<td>% Accuracy (correct/attempted)</td>
<td>87.4 (8.7)</td>
<td>84.2 (9.8)</td>
<td>0.34</td>
</tr>
<tr>
<td><strong>Double time (30 min)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Items correct</td>
<td>53.62 (10.02)</td>
<td>42.08 (1.51)</td>
<td>1.07</td>
</tr>
<tr>
<td>Items attempted</td>
<td>61.74 (9.60)</td>
<td>50.23 (12.49)</td>
<td>1.03</td>
</tr>
<tr>
<td>% Accuracy (correct/attempted)</td>
<td>86.8 (9.6)</td>
<td>83.8 (10.4)</td>
<td>0.29</td>
</tr>
</tbody>
</table>

### Table 2. Summary of Repeated Measures Analysis of Variance for Group and Time Conditions.

<table>
<thead>
<tr>
<th>Source</th>
<th>(df)</th>
<th>MSE</th>
<th>(F)</th>
<th>(p)</th>
<th>(\eta^2_p)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Items correct</td>
<td>1</td>
<td>4.457</td>
<td>20</td>
<td>&lt;0.01</td>
<td>0.213</td>
</tr>
<tr>
<td>Items attempted</td>
<td>1</td>
<td>4.365</td>
<td>18.9</td>
<td>&lt;0.01</td>
<td>0.204</td>
</tr>
<tr>
<td>Percentage correct</td>
<td>1</td>
<td>514.221</td>
<td>2.443</td>
<td>0.122</td>
<td>0.032</td>
</tr>
<tr>
<td><strong>Time Condition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Items correct</td>
<td>1</td>
<td>10.681</td>
<td>1.191</td>
<td>&lt;0.01</td>
<td>0.94</td>
</tr>
<tr>
<td>Items attempted</td>
<td>1</td>
<td>14.825</td>
<td>1.376</td>
<td>&lt;0.01</td>
<td>0.95</td>
</tr>
<tr>
<td>Percentage correct</td>
<td>1</td>
<td>3.754</td>
<td>0.414</td>
<td>0.661</td>
<td>0.006</td>
</tr>
<tr>
<td><strong>Group × Time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Items correct</td>
<td>1</td>
<td>87.0</td>
<td>9.7</td>
<td>&lt;0.01</td>
<td>0.116</td>
</tr>
<tr>
<td>Items attempted</td>
<td>1</td>
<td>95.0</td>
<td>8.88</td>
<td>&lt;0.01</td>
<td>0.107</td>
</tr>
<tr>
<td>Percentage correct</td>
<td>1</td>
<td>14.9</td>
<td>1.65</td>
<td>0.196</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Planned comparisons were conducted between the performance of students without disabilities at 15 min (standard time) and the performances of students with LD at 15, 22.5, and 30 min. LD students answered significantly fewer items correct than their nondisabled peers at 15 min, \(t(74) = -4.413, p < .01, d = -1.1\). However, when the LD group received time and one-half, they outperformed their standard-time peers (LD \(M = 31.04\), non-LD \(M = 26.38\)), \(t(74) = 2.6, p = .011, d = 0.61\). This difference increased further when only the LD group was given double time (LD \(M = 42.08\), non-LD \(M = 26.38\)), \(t(74) = 7.4, p < .001, d = 1.65\). These comparisons are presented in Table 3.
Time Condition showed that overall, participants attempted more items when given extended time, \( F(1, 148) = 1.376, p < .001, \eta^2_p = 0.95. \)

As with items correct, planned comparisons were conducted between the performance of students without disabilities at 15 min and the performances of students with LD at 15, 22.5, and 30 min. The LD students answered fewer items than their nondisabled peers at 15 min, \( t(74) = -4.148, p < .001, d = -1.03. \) When given time and a half, the LD group (\( M = 36.88 \)) outperformed their peers (\( M = 29.94 \)), \( t(74) = 3.64, p < .001, d = 0.84. \) And when given double time, students with LD (\( M = 50.23 \)) far outperformed their peers (\( M = 29.94 \)), \( t(74) = 9.0, p < .001, d = 2.0. \) These results also are summarized in Table 3.

**Table 3. Summary of \( t \) Tests.**

<table>
<thead>
<tr>
<th>Source</th>
<th>Source df</th>
<th>Mean difference</th>
<th>( t )</th>
<th>( p )</th>
<th>( d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LD.S vs. nondisabled.S</td>
<td>Items correct</td>
<td>74</td>
<td>-7.03</td>
<td>-4.41</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Items attempted</td>
<td>74</td>
<td>-6.79</td>
<td>-4.16</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>LD.H vs. nondisabled.S</td>
<td>Items correct</td>
<td>74</td>
<td>4.66</td>
<td>2.59</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>Items attempted</td>
<td>74</td>
<td>6.94</td>
<td>3.64</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>LD.D vs. nondisabled.S</td>
<td>Items correct</td>
<td>74</td>
<td>15.7</td>
<td>7.39</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Items attempted</td>
<td>74</td>
<td>20.3</td>
<td>9.06</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Note. \( S = \) standard time; \( H = \) time and a half; \( D = \) double time.

Discussion

**Interaction Hypothesis**

The interaction hypothesis predicted that the LD group would perform below nondisabled peers at standard time but catch up when both groups received extended time. Our results did not support this hypothesis. Indeed, there was an interaction in the opposite direction of the prediction: students without disabilities benefited more from extended time than their peers with LD. In other words, as the amount of extended time increased, the performance gap between students without disabilities and students with LD became even wider, favoring students without disabilities on both dependent measures of items correct and attempted. This finding suggests that extended time is not a test accommodation that is specific to those with LD, a violation of validity assumptions proposed by Phillips (1994).

As expected, non-LD students outperformed their peers with LD when the groups were compared at standard time. Of course, by definition, students with LD have difficulty with academic tasks such as reading, particularly when the reading task is time pressured. Our results, then, lend some confidence that the LD group, or at least some of its students, likely had genuine LD-related impairments.
Although the LD group answered fewer items at standard time, this changed dramatically when they alone were given 50% extra time. Some scholars have suggested that an extension between time and one-half and double time is usually enough to allow students with LD to complete tests, allowing them equal access (Ofiesh & Hughes, 2002). However, our results indicated that students with LD answered more questions at time and a half than students without disabilities did at standard time, and had a greater number correct. Whereas the nondisabled group had a performance advantage when both groups were compared at standard time, time and a half created a moderately sized advantage for the LD group when compared to the nondisabled group performance at standard time. Comparing students with LD at double time to students without disabilities at standard time, the LD group had a very large advantage in both items attempted and correct. Participants with LD had access to 26% more test content and had 20% more correct answers than typical students when only students with LD received double time. These findings suggest that extended time accommodations, particularly double time, go beyond “leveling the playing field,” and may reverse the playing field in favor of students with LD.

Validity of Test Accommodations

Results from the current study suggest that typical extended time accommodation practices may violate two of Phillips’ (1994) criteria of test accommodations’ appropriateness. First is the assumption of specificity, that an appropriate test accommodation should facilitate the performance of examinees with disabilities, but not nondisabled examinees. In the present study, students without LD actually performed increasingly better with more time. It is true that students with LD improved with extended time, but this gain was not specific to only them. The second violation is that the test scores of students given extended time, especially double time, seem to produce a very different outcome than standard time results, and this outcome strongly favors students with LD. It seems unlikely that an LD group should outperform a control group on a reading comprehension task, thus the double time scores may overestimate the skills of LD students.

The lack of specificity for extended time in the current study replicates the findings of previous research conducted by Lewandowski, Lovett, Parolin, Gordon, and Codding (2007) and Lewandowski et al. (2008). In each of these studies, typical students outperformed students with disabilities given extended time when ceiling effects were controlled. While the current study replicated the findings of previous research at 50% extended time, the current study also examined the effects of a double time condition. No previous study has systematically compared standard, 50%, and 100% extended time effects in college students with and without LD. The addition of a double time condition, which is a frequently granted test accommodation, extended the work of Lewandowski et al. It shows that as extended time increases to double time, there is even more evidence that this test accommodation lacks specificity. It is likely that the students without disabilities in this and other studies have the skills that allow them to make better use of the extended time.

Because the LD students answered fewer comprehension items at standard time, they may have warranted some additional time, if the primary goal of the accommodation is for students with LD to access the same number of items as students without disabilities. Interpolation from our data would suggest that 50% more time provides too much access, but that approximately 25% more time may have been just about right. This amount of extended time would have equated the number of items attempted by each group, a metric we could use to indicate degree of access to a test. Although this is only one study, the findings suggest that college students with LD may require some additional time on speeded tests to reach the same number of items as nondisabled examinees, but that 25% extra time may suffice for the typical LD
student, 50% extra would be more than what some students require, and 100% extra time would confer an unfair advantage for some students with LD. Of course, even this conservative guideline ignores that students without disabilities benefited from additional time as well.

Another validity criterion of Phillips’ (1994) was addressed in the study: Will the scores of the examinees tested under standard conditions have the same (or a different) meaning than scores for examinees tested with requested accommodations? Research in this area has been somewhat scarce. One prominent study by Cahalan et al. (2002) found that accommodated administrations of the SAT somewhat overpredicted GPAs of first-year college students with LD. A similar finding published by the Law School Admission Council (Thornton et al., 2001) noted that accommodated LSAT scores overpredicted first-year law school grades. While the current study utilized a different design than the above studies, we also found evidence that extended time results change the meaning of the test scores. At standard time, only 12% of our LD sample scored in the upper quartile of the test distribution for all students. However, with double time, the LD group placed 70% of its students in the top quartile of the distribution. Clearly, this is a dramatic shift in relative test outcomes for students with and without LD. Such a reversal could have significant implications for decisions based on a college admission test.

Limitations

The results of the current study are tempered by several limitations. First, the LD students were drawn from two private, competitive postsecondary institutions. Despite their disability diagnoses, these students would have had to attain a respectable high school GPA and standardized test scores to meet the institutions’ minimum requirements. As such, it may be that the LD sample in this study is less impaired than the LD population in general. Recent research by Sparks and Lovett (2009) suggests that many students with LD diagnoses in college populations may not have significantly impaired academic achievement, compared to the general population. In their review, most college LD students showed average achievement scores, but lower scores than other college students. In similar fashion, the LD sample in this study may not be substantially impaired in academic achievement. We can only note that they had lower mean GPAs, had professional diagnoses, and were all receiving test accommodations in college. Therefore, our results may lack generalization to the broader population of all LD students.

Another limitation of the study is a lack of control of comorbid conditions. For example, 50% of the LD sample (n = 13) reported a comorbid condition such as ADHD, depression, or anxiety. Whereas comorbid conditions in the LD population are quite common, there is still the question of whether the student is impaired just by the LD or a combination of difficulties. In this study, students with comorbid diagnoses had slightly lower performance scores although the differences were not statistically different (power was low in these analyses). Future researchers may wish to use LD samples without comorbid disorders, or else broaden their hypotheses to compare groups of students with other disabilities (i.e., ADHD, anxiety).

Third, although our primary measure, the modified NDRT comprehension test, was carefully constructed to minimize both ceiling effects and examinee fatigue, our results may not generalize to other kinds of tests. It is possible that on different tests of reading fluency and comprehension the students without LD would not benefit as much from extended time as the students with LD. More specifically, students without disabilities were generally unable to finish our test before the “standard” time limit was up, and our results are unlikely to generalize to tests with very liberal standard time limits. Of course, on many high-stakes tests (e.g., the SAT), students without disabilities report being unable to finish within the standard time limits, and so our results may generalize well to these kinds of situations.
Finally, the current study used a reading comprehension measure that is similar to sections found on high-stakes exams such as the SAT, ACT, LSAT, and GRE, but we could not simulate the high-stakes test atmosphere. However, it appeared that both groups nonetheless demonstrated reasonable effort. Students in both groups answered a reasonable number of questions and performed with a high degree of accuracy on the test items (LD = 83% and nondisabled = 88% accuracy). Also, it is noteworthy that accuracy scores were similar not only for both groups but also for each time condition. This suggests that students did not lose interest or effort in the task as time increased. Even though this was not a high-stakes test setting, it seems that students provided reasonable and consistent effort.

Conclusion
The results of the current study certainly question the appropriateness of lengthy time extensions as appropriate accommodations for college students with LD, even though all of these students were receiving extended time. Our data support prior work showing that extended time benefits students without disabilities and is not of specific benefit to those with a disability. Furthermore, we found that the amount of time given to a student on a speeded exam can make a dramatic difference in one’s performance outcome relative to other students. It is no wonder that extended time is highly desired by students that apply for test accommodations, whether their problem involves reading, spelling, anxiety, attention, or general slowness.

In the absence of support for extended time as a disability-specific test accommodation, test makers should consider “universal test design” approaches that would be fair and valid for all students (Lewandowski et al., 2008; Lovett, 2010; Royer & Randall, 2011). Universal test design means that a procedure is developed that gives all students the opportunity for equal access to the exam so they can display their full skill levels. To effectively use universal design, test developers, administrators, and users must determine what constructs are important foci of measurement. It is time for testing entities to initiate research on the importance (or lack thereof) of speed in performance; if speed is truly of no importance, time limits should be liberalized for all examinees, and if speed has any relevance, time limits should be empirically determined. Extended time accommodations might only be given in special situations, therefore avoiding some of the rampant and arbitrary use we see today.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

References


