

## ARTICLES

---

# Effects of a Student-Reads-Aloud Accommodation on the Performance of Students With and Without Learning Disabilities on a Test of Reading Comprehension

Batya Elbaum, Maria Elena Arguelles, Yvonne Campbell,  
and Maya Bardawil Saleh

*Department of Teaching and Learning  
School of Education  
University of Miami*

In this study, we investigated the impact of a student-reads-aloud accommodation on the performance of middle school and high school students with and without learning disabilities (LD) on a test of reading comprehension. Data for the analyses came from 311 students ( $n = 230$  with LD) who took alternate forms of a reading test in a standard and an accommodated condition. In the accommodated condition, students were instructed to read each passage aloud at their own pace and then to read each comprehension question and the response choices aloud before marking their answer. As a group, students' test performance did not differ in the 2 conditions, and students with LD did not benefit more from the accommodation than students without LD. However, students with LD showed greater variability in their response to the accommodation such that they were almost twice as likely as students without LD to show a substantive change in test performance in either the positive or negative direction. The findings of the study underscore the need to go beyond the interpretation of group mean differences in determining the validity of testing accommodations.

In recent years, the increasing participation of students with disabilities in statewide assessments has stimulated considerable research and discussion concerning the appropri-

ate assignment of testing accommodations, the impact of accommodations on test performance of students with and without disabilities, and the validity of interpretations of test performance when students are awarded particular accommodations (Elliott, McKeivitt, & Kettler, 2002; Thurlow & Bolt, 2001; Thurlow, House, Scott, & Ysseldyke, 2000; Thurlow, McGrew, et al., 2000; Tindal, 2002; Tindal & Fuchs, 1999). Issues surrounding testing accommodations have important implications both for accountability systems as well as for individual students (Elliott & Roach, 2002; Individuals with Disabilities Education Act, 1997; No Child Left Behind Act, 2001; Tindal, 2002). Accountability systems must take into consideration whether test scores of students who are awarded accommodations can be considered commensurate with scores of students who take the tests without accommodations. For individual students, the appropriate assignment of accommodations for high-stakes tests could make the difference between passing to the next grade and retention or between exiting school with or without a standard diploma.

The urgency of the issue has led to a burgeoning research literature on testing accommodations for students with disabilities. To date, reviews of the testing accommodations research (e.g., Chiu & Pearson, 1999; Thurlow & Bolt, 2001; Tindal, 2002) indicate that considered for groups of students, the effects of accommodations on test performance are generally quite small. A meta-analysis by Chiu and Pearson (1999, as cited in Tindal, 2002) found that studies using general education students as a comparison group yielded an overall weighted mean accommodation effect size for all target population students of 0.16, with a standard error of 0.02. The synthesis also revealed large and statistically significant variation in the effects associated with different accommodations, supporting the need to understand the effects associated with specific accommodations. Tindal and others have noted the complex nature of accommodation effects, underscoring the importance of investigating possible interactions between the accommodation effect, student disability, student skill level in the area being tested, and characteristics not only of tests but also of specific test items.

A large percentage of students with disabilities have severe difficulty in reading and are candidates for testing accommodations on tests of reading comprehension. Students with learning disabilities (LD) make up almost 50% of all students with disabilities, and the vast majority of these students have individualized educational program goals in reading. This is the population of students with disabilities we investigated in this study.

Several studies have investigated the impact of testing accommodation on the performance of students with LD on tests of reading comprehension. Fuchs et al. (2000) administered a reading assessment to fourth- and fifth-grade students with LD and fourth graders without LD under four different testing conditions: standard, large print, extended time, and student reads aloud. Students did not benefit from extended time or large print. Marquart (2000, as cited in Elliott et al., 2002) similarly found no statistically reliable effect for an extended time accommodation on reading tests. In contrast, students with LD in the Fuchs et al. study benefited significantly more than students without LD from reading the passages aloud. For the student-reads-aloud accommodation, there was a significant difference in the accommodation effect for students with LD (effect size [ES] = 0.06) and students without LD (ES = -0.12).

Thus, of the various accommodations for reading tests that have been investigated, only one—allowing the student to read the passages aloud—has been found to produce a differential gain in the performance of students with LD. Reasons to explain this pattern of results have varied. When extended time fails to enhance scores of students with LD, it may be because their knowledge and skills in a particular area may not be commensurate to the difficulty of the test. When extended time enhances the performance of all students, not only those with LD, the accommodation is not considered valid. The theoretical implication is that if time affects all students equally, then all students should take the test under the same time conditions or else students taking the test in the accommodated condition would actually have an unfair advantage over other students.

The explanation of absence of impact for large print is more straightforward. For students with no visual impairment, there is no benefit to having passages displayed in a larger font. This is equally true of students with and without reading disabilities.

With regard to students reading aloud, Goldman, Hogaboam, Bell, and Perfetti (1980) studied elementary school students' recall of specific words read within a sentence and across a sentence boundary. They divided their sample of students into those of higher and lower reading ability based on teacher reports and had students read the stimulus material in one of two conditions: silently or aloud. They found that students of lower reading ability, particularly younger students (third vs. fourth graders), had greater recall for text just processed when they read the text aloud. Although this finding was somewhat incidental to their main investigation, it fit with the view of reading comprehension as being dependent on holding just-read text in short-term memory until sufficient text (usually a clause) has been processed to encode a complete meaning unit. It would follow that if reading aloud assists less highly skilled readers to recall specific text long enough to enhance comprehension, then allowing students to read the passages of a reading test aloud might constitute an appropriate testing accommodation for students with LD. Based on the evidence provided by Goldman et al., this accommodation would not be likely to benefit more highly skilled readers and thus would meet the criterion for a valid testing accommodation. This is in fact what Fuchs et al. (2000) found to be the case. However, based on the Goldman et al. findings, it is unclear whether much older students (i.e., those in middle school and high school) would reap the same advantage.

Several researchers have also investigated teachers' and/or students' perceptions of the impact of specific testing accommodations on student performance. Fuchs et al. (2000) found that teachers were not accurate in their assignment of testing accommodations; they awarded accommodations to students who did not benefit from them and did not award them to students who did benefit from them. Helwig and Tindal (2003) similarly found that teachers were not effective in predicting who would benefit from an accommodation. McKeivitt and Elliott (2003) reported that eighth-grade students, responding to a questionnaire regarding test accommodations, thought they did better on tests when accommodations were provided. However, no analyses were conducted to verify whether the degree to which students perceived the accommodation to be effective was associated with the accommodation boost they experienced.

Thus, this study was designed to accomplish two primary goals. First, extending the work of Fuchs et al. (2000) to an older group of students, we wished to study the impact of the student-reads-aloud accommodation on the reading test performance of students

with and without disabilities in middle school and high school. Second, we wished to examine the accuracy of students' perceptions of the impact of this accommodation on their test performance in reading.

## METHOD

### Participants

Participants in the study were 456 students (283 with LD; 276 male) in Grades 6 through 10. The students were recruited from six schools (three middle schools and three high schools) in a large urban school district in the southeastern United States. The school population in the district is highly diverse in terms of race, ethnicity, and socioeconomic status. Table 1 presents the distribution of participating students by grade grouping and gender.

### Measures

The reading tests used in this study were constructed using third- to fifth-grade level reading passages and accompanying comprehension questions designed for use as test preparation exercises in language arts classes. We ascertained in advance that the specific passages being utilized had not been included in any practice activities at the participating schools.

We initially administered on-grade-level passages to a sample of students with LD in Grades 7 and 9 attending schools that were comparable to schools participating in the study. The passages and test questions were similar in content, presentation format, and response format to those on the statewide reading assessment. The purpose of this pilot test was to ascertain whether the target students' performance in a standard administration condition was adequate for our study. That is, test passages that produced a floor or ceiling effect would not yield accurate information on the potential benefits of an accommodation. The distribution of students' scores on a set of grade-level passages was in fact highly positively skewed, with many students unable to answer more than a few test items correctly. Given this outcome, we tested the students on a selection of third-through fifth-grade level passages. The distribution of students' scores using these pas-

TABLE 1  
Distribution of All Participating Students by Grade Grouping and Gender

	<i>Students With LD</i>		<i>Students Without LD</i>	
	<i>Male</i>	<i>Female</i>	<i>Male</i>	<i>Female</i>
Middle school (Grades 6–8)	114	59	45	51
High school (Grades 9–10)	68	42	49	28
Total	182	101	94	79

*Note.* LD = learning disabilities.

sages was approximately normal, with mean performance around 50%. Consequently, we determined that although our state does not permit off-grade-level testing on the statewide assessment, for purposes of detecting an accommodation effect, we would have to use easier reading passages than those that the students would actually confront on the end-of-year statewide assessment.

To create two alternate test forms of equal difficulty, we administered a test consisting of 9 third- through fifth-grade reading passages of varying length and genre to another sample of students. Based on students' responses to test items, the items were scaled from easiest to most difficult and the passages ranked in terms of the average number of items to which the pilot sample of students responded correctly. Based on the difficulty order of the passages, we eliminated one passage that was markedly more difficult for students than the other eight. From the remaining eight, we created two test forms balanced for overall item difficulty and passage length.

To test the equivalence of the two alternate forms, we administered both forms to a sample of 98 students (24 middle school students with LD, 40 middle school students without LD, and 34 high school students with LD). Students' performance on the two forms showed a correlation of  $r = .83$ . (Calculated separately for the three subgroups listed previously, the correlations were  $r = .73$ ,  $r = .89$ , and  $r = .79$ , respectively.) We deemed this result to be adequate for our purposes.

*Individual interview.* Each student was individually interviewed after participating in the two test conditions. The interview took less than 5 min. We asked students how difficult or easy they perceived the test to be and whether they felt they had performed better in the silent reading condition, performed better in the read-aloud condition, or performed about the same in the two conditions.

## Procedure

*Recruitment.* Once district, school, and institutional review board approval had been obtained to conduct the study, a member of the research team went to multiple language arts classrooms in each of the six schools and explained the purpose of the study. Students were given a letter and a parental consent form to take home and were asked to return the form regardless of the decision they and their parents made concerning their participation. Participating students also signed a student assent form.

*Administration of reading assessments.* Test forms but not test conditions were counterbalanced across intact classrooms. For all students, we first administered the test in the standard (read silently) condition. Students who had taken the test in the standard condition were then scheduled for individual testing in the accommodated condition. Students took the accommodated test between 2 and 3 weeks following testing in the standard condition, a period we felt was long enough to hold any practice effect to a minimum. (If a practice effect were, in fact, present, it would result in a more conservative estimate of the accommodation effect.)

The tests were administered by one of the authors or by a trained research assistant. In the standard condition, the test was group administered to students in a regular classroom

during their language arts period. Prior to beginning the test, a sample test item was reviewed to ensure that students understood the test format. In the standard condition, the test administrator signaled the students when to begin and end each reading passage and set of test questions. To ensure consistent implementation of the test procedure, the testers followed a written test administration script and used a stopwatch.

The time allotted for silent reading of the passages was determined by application of an algorithm based on the procedure used by Fuchs et al. (2000). The passages used by Fuchs et al. were very similar to one another in length, and the allotted time of 2 min per passage represented an approximate time of 3 sec per word. Because our passages were much more variable in length, we multiplied the word length of each passage by 3 sec and rounded to the nearest half minute. Similar to Fuchs et al., we gave students 30 sec per question (total time = 2 min) to answer the questions following each passage. Total administration time for the group-administered standard condition, including instructions, was approximately 25 min.

In the student-reads-aloud condition, the test was individually administered in the school library. Students were instructed to read the passage aloud at their own pace, then to read the test items aloud and mark their answers. The test administrator explained that the accommodation was intended to support the student's performance, that the student was not being evaluated on oral reading, and that the administrator would not provide any feedback or corrections. The test administrator told the student to begin the test and then sat down at a different table so that the student did not feel that he or she was being listened to. The administrator intervened to prompt the student only if the student appeared not to be reading the passages aloud.

*Exit interview.* The individual interview was conducted at the conclusion of the individually administered accommodated test condition. At the end of the interview, students were given an opportunity to ask any questions they had about the study. Students were thanked for their participation and given a small gift (e.g., university logo pencil) in appreciation.

## RESULTS

Test means and standard deviations of students with and without LD in both conditions are reported in Table 2. For all students combined, the distribution of scores had a mean of 10.7 and standard deviation of 3.5. The distribution showed significant negative skew owing to many students achieving the maximum or near-maximum score.

We conducted our main analyses on a subset of these students defined as those who achieved scores between 5 and 13 (out of a maximum score of 16) in the standard condition. The reason for excluding students who performed within 2 points of the maximum score is that these students were already achieving such high scores that it would be impossible to detect improvement in the accommodated condition. Calculating an estimated accommodation effect based on such a high-performing sample could yield a highly biased estimate in the negative (null) direction.

The reason for excluding students who scored 4 or below is that scores this low could have been obtained by chance alone. Thus, the interpretation of scores in this range is highly ambiguous. Students who score at or below chance may in fact have reading skills that are so far below the level necessary to respond authentically to the test items that no type of test accommodation would close the gap. Including students who perform no better than chance in the unaccommodated condition could bias against a positive finding of accommodation effects, not because the accommodation itself is ineffective but because of the extreme mismatch between the difficulty of the test and the skill level of the students.

Given our very large sample size, we opted to trim our sample so that insofar as possible, we were testing the effect of the accommodation on students for whom the test was moderately challenging, that is, neither so difficult that they could not perform above chance nor so easy that they could get almost all the items correct. In this sense, students with and without LD were roughly equated on their performance in the standard condition.

Truncating the distribution resulted in the exclusion of 145 of the 456 students who participated in testing (32%). A total of 53 of 283 students with LD (19%) were excluded, 27 because they performed at or below chance and 26 because they obtained scores of 14 or above. Also excluded were 92 of 173 general education students (53%), all of whom obtained scores of 14 or above.

Students remaining in the analysis were 230 students with LD and 81 students without LD (see Table 3 for the distribution of these students by grade grouping and gender). Table 4 presents the test means and standard deviations for this subset of students.

TABLE 2  
Test Performance of Students With and Without LD in Each Test Condition

Condition	Students With LD <sup>a</sup>		Students Without LD <sup>b</sup>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Standard	9.19	3.36	13.20	2.30
Student reads aloud	9.21	3.50	12.62	2.17

Note. LD = learning disabilities.

<sup>a</sup>*n* = 283. <sup>b</sup>*n* = 173.

TABLE 3  
Distribution of Students in Main Analyses by Grade Grouping and Gender

	Students With LD		Students Without LD	
	Male	Female	Male	Female
Middle school (Grades 6–8)	91	47	16	24
High school (Grades 9–10)	54	38	23	18
Total	145	85	39	42

Note. LD = learning disabilities.

TABLE 4  
Test Performance of Students With and Without LD in Each Test Condition

Condition	Students With LD <sup>a</sup>		Students Without LD <sup>b</sup>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Standard	9.27	2.45	9.16	3.25
Student reads aloud	11.52	1.70	11.86	2.39

Note. Students whose data are reported in this table were those who achieved scores between 5 and 13 (out of a maximum score of 16) in the standard test condition. LD = learning disabilities.

<sup>a</sup>*n* = 230. <sup>b</sup>*n* = 81.

### Overall Effects

To investigate the overall effects of the accommodation, a  $2 \times 2$  repeated measures analysis of variance (ANOVA) was conducted with test condition (standard vs. accommodated) as a within-subjects variable and disability status (LD vs. non-LD) as a between-subjects variable. The between-subjects effect was statistically significant,  $F(1, 309) = 67.54, p < .001$ , indicating that students without LD performed significantly better than students with LD. In contrast, neither the main effect for test accommodation,  $F(1, 309) = 0.49, p = .49$ , nor the Disability  $\times$  Accommodation interaction effect,  $F(1, 309) = 1.69, p = .19$ , was statistically significant.

Because previous research has reported accommodation results in terms of the accommodation boost—a term that connotes a positive change in performance resulting from use of an accommodation but that can actually be either positive, negative, or zero—we calculated the boost, as in previous research (e.g., Fuchs et al., 2000), as the difference between a student's score in the accommodated condition and the student's score in the unaccommodated condition. For both groups together, the accommodation boost ranged from  $-7$  to  $7$ , with a mean of  $0.01$  and a standard deviation of  $2.70$ . For the two groups separately, the distribution values were as follows: For students with LD,  $M = -0.10, SD = 2.81, \text{range} = -7 \text{ to } 7$ ; for students without LD,  $M = 0.35, SD = 2.24, \text{range} = -5 \text{ to } 6$ . Calculated as a within-group effect size (mean group score in accommodated condition minus mean score in standard condition divided by the standard deviation of scores in the standard condition), the accommodation effect for students with LD was  $-0.10/2.45 = -0.04$ . The effect size for students without LD was  $0.35/1.70 = 0.21$ . Recapitulating the findings of the ANOVA, the difference in accommodation boost between students with and without LD was not statistically significant,  $t(309) = -1.30, p = .19$ .

Given the differential performance of students with and without LD in the standard condition, the foregoing analyses may not have fully controlled for statistical regression, which exists even in a truncated sample. As a group, students without LD performed above the mean, whereas students with LD performed below. Thus, on a repeated measure, the scores of students with LD would be predicted to be higher (closer to the mean) and those of students without LD would be predicted to be lower (closer to the mean). The accommodation boost was thus recalculated as the residualized change score; that is, the gain was calculated with regression effects removed (Campbell & Kenny, 1999). The



mean residualized change scores for the two groups of students were  $-0.35$  for students with LD and  $1.01$  for students without LD. A  $t$  test for the difference between these means was statistically significant,  $t(309) = -4.05$ ,  $p < .001$ . Thus, when regression effects were more stringently controlled for, students without LD, as a group, appeared to have benefited more from the accommodation than students with LD.

### Odds of Benefit or Detriment

We also analyzed the data in terms of the odds that a given student would benefit from the test accommodation under investigation. This analysis reflects the admittedly not ideal, but typical, circumstance in which teachers have no prior basis for determining whether or not to assign a particular accommodation to a given student.

We defined three categories of response to the accommodation based on the standard deviation of scores in the unaccommodated condition ( $SD = 2.48$ ). If students' scores in the accommodated condition were 3 or more points higher than in the standard condition ( $> 1 SD$  unit gain), we defined these students as having benefited from the accommodation. For a student performing at the 50th percentile, a gain of one standard deviation unit would result in a performance at the 84th percentile. Analogously, students whose scores were 3 or more points lower in the accommodated condition were identified as having suffered a detriment as a result of the accommodation. The cross-tabulation of accommodation benefit by disability status is displayed in Table 5.

As seen in the table, the performance of students with LD was more likely than that of general education students to be impacted, for good or for ill, by the accommodation. The percentage of students who registered substantive gain or loss was 37% for students with LD compared to 20% for students without LD. For a student with LD, the odds of benefiting from the read-aloud accommodation were 1 in 5 (20%); for a general education student, the odds were approximately 1 in 10 (11.1%). Similarly, for a student with LD, the odds of obtaining a lower score in the accommodated condition were almost 1 in 6 (17.4%); for a general education student, the odds were only 1 in 10 (9.9%). Thus, students with LD appear to have double the likelihood of benefit and almost double the likelihood of detriment from blanket assignment of this accommodation.

Given the approximately equal distribution of individual values of the accommodation boost around a mean close to zero, we conducted an additional analysis that we felt might shed light on the question of whether or not, overall, the accommodation made a difference in students' scores. Specifically, we compared the correlation of scores across test conditions with the correlation of scores obtained when we administered both forms of the reading assessment to students in a standard condition. As reported earlier, the correlation between forms for all three groups of pilot students ( $n = 98$ ) was  $r = .83$ . In contrast, the correlation between students' scores under two different test conditions ( $n = 311$ ) was  $r = .60$ . Using formulas and tables provided by Hays (1981), we calculated the test statistic for a difference between correlations for independent samples. The obtained value was highly statistically significant,  $z = 4.21$  (a value of 1.96 would be significant at  $p = .05$ ). The statistically significant statistic suggests that an effect other than random error was operating in the accommodated test condition even though the effect did not produce the same result (i.e., increased scores) for all students.

TABLE 5  
Cross-Tabulation of Accommodation Benefit by Disability Status

<i>Disability Status</i>	<i>Accommodation Benefit</i>			<i>Total</i>
	<i>Benefit</i>	<i>No Difference</i>	<i>Detriment</i>	
<b>Students with LD</b>				
Count	40	144	46	230
% within disability status	17.4	62.6	20.0	100.0
% within accommodation benefit	83.3	69.2	83.6	74.0
% of total	12.9	46.3	14.8	74.0
<b>Students without LD</b>				
Count	8	64	9	81
% within disability status	9.9	79.0	11.1	100.0
% within accommodation benefit	16.7	30.8	16.4	26.0
% of total	2.6	20.6	2.9	26.0
<b>All students</b>				
Count	48	208	55	311
% within disability status	15.4	66.9	17.7	100.0
% within accommodation benefit	100.0	100.0	100.0	100.0
% of total	15.4	66.9	17.7	100.0

*Note.* LD = learning disabilities.

### Students' Self-Perceptions

Data from the individual interview were available for 302 of the 311 students whose scores were used in the main analyses. Fewer than half the students with LD ( $n = 92$ ; 41%) perceived that they had performed better in the accommodated condition, 81 (36%) perceived that their performance was better in the silent reading condition, and 51 (23%) perceived that their performance in the two conditions was about the same. The pattern for students without LD was roughly similar. Of these students, 38 (49%) perceived that their test performance was better in the accommodated condition, 15 (19%) perceived that their performance was better in the silent reading condition, and 25 (32%) perceived that their performance in the two conditions was about the same.

To investigate the accuracy of students' perceptions, we cross-tabulated students' expressed perception of benefit (better performance in the student-reads-aloud condition, worse performance in the student-reads-aloud condition, or no difference) with their actual benefit from the accommodation (benefit, detriment, or no difference). The analysis is found in Table 6. The upper-left to lower-right diagonal in the table represents accurate perceptions; the off-diagonal cells represent inaccurate perceptions. The value of chi-square,  $\chi^2(4, N = 302) = 9.84, p = .04$ , was statistically significant, indicating that students were more accurate than chance.

However, to say that accuracy was slightly better than chance is not to say that it was good. Of those who benefited from the accommodation, 56.4% reported (correctly) that the accommodation had improved their performance, 12.8% reported no difference, and 30.8% reported (erroneously) that the accommodation hindered their performance. Of those whose performance was negatively affected by the accommodation, 31.1% re-

TABLE 6  
Cross-Tabulation of Actual and Perceived Accommodation Benefit

<i>Perceived Accommodation Benefit</i>	<i>Actual Accommodation Benefit</i>			<i>Total</i>
	<i>Benefit</i>	<i>No Difference</i>	<i>Detriment</i>	
<b>Benefit</b>				
Count	28	85	17	130
% within perceived benefit	21.5	65.4	13.1	100.0
% within actual benefit	59.6	42.3	31.5	43.0
% of total	9.3	28.1	5.6	43.0
<b>No difference</b>				
Count	6	55	15	76
% within perceived benefit	7.9	72.4	19.7	100.0
% within actual benefit	12.8	27.4	27.8	25.2
% of total	2.0	18.2	5.0	25.2
<b>Detriment</b>				
Count	13	61	22	96
% within perceived benefit	13.5	63.5	22.9	100.0
% within actual benefit	27.7	30.3	40.7	31.8
% of total	4.3	20.2	7.3	31.8
<b>Total</b>				
Count	47	201	54	302
% within perceived benefit	15.6	66.6	17.9	100.0
% within actual benefit	100.0	100.0	100.0	100.0
% of total	15.6	66.6	17.9	100.0

ported (incorrectly) that the accommodation helped them, 22.2% reported no difference, and 46.7% reported (correctly) that the accommodation impeded their performance. Indeed, when the data were cross-tabulated separately for students with and without LD, the groups' perceptions were not statistically more accurate than would have occurred by chance alone.

## DISCUSSION

This study examined the effects of a student-reads-aloud accommodation on the performance of middle school and high school students with and without LD on a test of reading comprehension. Overall, the test scores that students achieved in the accommodated condition were not statistically significantly different from scores obtained in the standard condition. For 17% of students with LD, the accommodation boosted performance; for 20%, the accommodation impaired performance. Considering students without LD, 10% showed an accommodation benefit, whereas 11% showed an accommodation detriment. Although an ANOVA revealed no statistically reliable difference in the accommodation boost for students with and without LD, the analysis of residualized gain scores suggested that at approximately equal levels of performance in reading comprehension, students without LD may, as a group, benefit more from this accommodation than students with LD.

This study implemented several recommendations in the literature concerning the experimental investigation of testing accommodations. Tindal, Heath, Hollenbeck, Almond, and Harniss (1998) urged that "to provide the most convincing empirical support for an accommodation, students with a specific need have to be compared to others without such a need who are otherwise comparable in achievement" (p. 442). In this study, a comparison group of general education students was chosen that was very close to the reading performance level of the students with LD. As well, students took the test in both conditions, thus acting as their own controls.

The finding that as a group, students with LD did not have higher scores in the accommodated than in the standard test condition calls into question the efficacy of the accommodation; the fact that they did not benefit more from the accommodation than students without LD calls into question its validity. With regard to efficacy, it may be the case that as suggested by Goldman et al. (1980), older students with low reading skills are less likely than younger ones to benefit from producing an overt phonological representation of the text. Although reading aloud may enhance retention of discourse in working memory, it also slows down reading speed. For older readers, the trade-off of increased retention versus slower processing of the text, particularly for longer passages, may not be sufficiently advantageous to result in overall gains in comprehension.

With regard to validity, the findings of our study raise the question of whether scores achieved in the accommodated condition can be interpreted in the same way as scores obtained in the unaccommodated condition. Tindal (2002) described several perspectives on the validity of accommodations. For example, Phillips (1994) specified five conditions for an appropriate definition of accommodations including that the meaning of scores should be the same regardless of any changes being made in the manner in which the test is given or taken and that the accommodation should not have the potential for benefit for students without disabilities (for further discussion, see Elliott et al., 2002). The implication is that a necessary condition for test validity under accommodated conditions is that students without disabilities do not benefit from the accommodation (cf. Tindal et al., 1998). In the case of Fuchs et al. (2000), the student-reads-aloud accommodation was concluded to be valid because although the gain for students with LD in the student-reads-aloud condition was very small ( $ES = 0.06$ ), students without LD suffered a detriment in performance ( $ES = -0.12$ ), resulting in a significant differential accommodation boost.

However, various researchers, including Fuchs et al. (2000), have cautioned against the group differences approach to defining validity. Fuchs et al. asserted the following:

Mean group differences represent only one yardstick by which the validity of test accommodations should be assessed. Along any dimension, whenever group averages differ, the populations inevitably overlap. This means that, despite the group patterns, some individuals with LD will profit differentially from extended time or large print; others will fail to profit in important ways from reading aloud. This is why decisions must be formulated individually to limit accommodations to the subset of students who realize greater boosts than expected for students without LD. (pp. 76-78)

Elliott and Roach (2002) pointed out:

The lack of a differential benefit alone may not be sufficient to conclude invalidity of scores resulting from the use of accommodations. . . . The accommodations still may have served to remove a disability-related barrier for the student tested, yet still did not have a significant effect on scores. Thus evidence to support the validity of accommodations needs to come from multiple sources. (p. 17)

Moving beyond mean group differences in accommodation boost, we investigated alternative evidence of an accommodation impact in this study. We interpreted the lower correlation of scores across testing accommodations compared to the correlation across alternate test forms in the standard condition to indicate that the accommodation did have an impact. Of potential interest is the fact that students with LD showed greater dispersion of the accommodation boost than general education students such that the students with LD were twice as likely to be substantially impacted by the accommodation. In effect, the performance of students with LD was more likely to be “perturbed” in one direction or the other as a consequence of the accommodation. This is in line with evidence regarding the considerable heterogeneity of students with LD (Morris et al., 1998). The finding is also consistent with findings of McKeivitt and Elliott (2003, as cited in Elliott et al., 2002). McKeivitt and Elliott studied the impact of teacher-recommended accommodations on the performance of students with and without disabilities on a reading test. They found considerable variability in the accommodation effects; the accommodations positively affected the scores for half of all students with disabilities and 38% of all students without disabilities. There is no report of any negative impact on students’ performance as a result of the accommodations.

It is important to note that the foregoing discussion of validity is predicated on the premise that accommodations have the potential to alter the construct being tested and that it must be demonstrated that they do not do so. With regard to the accommodation investigated in this study—having the student read reading passages aloud—it could be asserted that the construct of reading is indifferent to whether or not a text is read silently or aloud. Although a full discussion of the controversy surrounding models of skilled reading would go beyond the purpose of this study, several comments are relevant to an understanding of the accommodation we have investigated.

In our view, proponents of a strong phonological theory of visual word recognition (e.g., Frost, 1998) present compelling arguments against the assertion that skilled performance in reading involves bypassing the mechanisms that convert orthographic structures into phonological structures. A more parsimonious explanation of skilled reading is that with practice, the reader’s efficiency in computing a prelexical phonological representation increases; the reader also acquires greater efficiency in accessing the lexicon with impoverished phonological information. In this view, beginning readers or older readers who have not acquired a high level of efficiency in reading must undertake a detailed phonological analysis of the printed word before lexical access, and hence, comprehension is achieved. Poor readers or at least those whose primary difficulty lies in

phonological decoding may benefit from reading aloud because it helps them arrive at the more complete phonological analysis they need to achieve lexical access.

If the construct of reading comprehension is not constrained to acts of silent reading, then we must reevaluate both the premise of the accommodation and the experimental findings. The benefit of the accommodation—as in the Goldman et al. (1980) study that suggested it—may accrue to any lower skilled reader whether or not the lower level of skill is due to a specific disability. Indeed, from this perspective, the accommodation is only an accommodation to conventional test administration practices, which for the sake of efficiency typically involve the testing of large groups of students in the same room at the same time. In essence, there is no theoretical reason why all students should not have the option of reading a reading test aloud. Certainly, some students subvocalize as they read and are permitted to do so as long as they do not do so (very) audibly. The finding that students without disabilities responded similarly to students with LD, albeit less extremely, can be viewed as supporting the use of the accommodation with all students who might benefit from it, irrespective of disability status.

On a separate issue, our findings also underscore the importance of taking an individual perspective on testing accommodations and of requiring that accommodation decisions be based on trials undertaken by each student. In most test accommodation studies, students' scores either increased as a result of the accommodation or remained unchanged. In this study, in contrast, some students suffered a potential harm (impaired test performance) as a consequence of the accommodation. The potential for harm makes it essential that great caution be applied in assigning this accommodation and that it be assigned only on the basis of prior evidence of benefit to the individual student.

With regard to students' perceptions, previous research indicated that students are generally well disposed to testing accommodations. For example, McKeivitt and Elliott (2003) found that students in their study had positive views of a tester-reads-aloud accommodation, although they expressed some concern that having the test items read aloud made them difficult to follow. Elliott et al. (2002) reported on a dissertation study by Marquart (2000) investigating extended time on a mathematics test for eighth-grade students. Students in this study were surveyed concerning their perceptions of the accommodation. Most reported that they felt more comfortable, were more motivated, thought they had performed better, and preferred taking the test with the extended time accommodation. Interestingly, neither the effect size for students without disabilities ( $ES = 0.34$ ) or that for students with disabilities ( $ES = 0.26$ ) was statistically different from zero.

Our study indicated that students were not very accurate in their perceptions of whether or not their test performance was enhanced as a result of the accommodation. In contrast to previous research investigating teachers and students' predictions of benefit, this study examined students' postdictions of benefit. That is, students in this study had the potential advantage of having experienced the accommodated condition just prior to being queried about their perceptions. Still, the postdictions of students in this study, although statistically slightly better than chance, were far from accurate. Translated into practical application, the experience of the accommodation in and of itself did not provide students with an accurate basis for determining whether they would be appropriate candidates for this accommodation. Perhaps, with repeated experience and feedback on results, their accuracy would improve. The inability to accurately assess the impact of the

accommodation characterized students without disabilities as well as students with LD. Thus, students appear to be no more accurate than teachers (cf. Fuchs et al., 2000; Helwig & Tindal, 2003) in their perceptions of the actual or potential impact of an accommodation on their test performance.

### Limitations

A serious limitation of many studies of testing accommodations including this one is the confounding of the accommodation (in our case, students reading the text aloud) with concomitant factors such as self-pacing and individual administration. Although the time allotted on the group administration was fairly generous—it was noted that almost all students appeared to finish all the passages—some students may not have had time to fully process the text. Moreover, some students in the silent condition may not have read through the passage in its entirety, a scenario that appears less likely when a student is reading aloud.

Authors of previous accommodation studies have pointed to similar confounds in their work. For example, Tindal et al. (1998) provided a tester-read-aloud testing accommodation to students with and without an LD in reading on tests of science, usage and expression, and math problem solving and data interpretation. According to Tindal et al., the tester-read-aloud test administration format did not permit student self-pacing, which could have affected students' maintaining attention to the test. Similarly, Hollenbeck, Rozek-Tedesco, Tindal, and Glasgow (2000, as cited in Tindal, 2002) underscored that in their accommodation study, which utilized a teacher-read-aloud accommodation on a mathematics test, the accommodation actually involved both the read-aloud and extended time. Thus, it is possible that in this study, students' performance in the accommodated condition reflected not only the effect of the student reading aloud but also the effect of more relaxed timing and the more private setting. Unfortunately, the design of our study did not allow us to partial out the contributions of these various components.

### Implications for Research

The findings of this study lend additional support to the recommendations of Elliott et al. (2002), Tindal (2002), Tindal et al. (1998), and others with regard to future accommodations research. First, the effects of different components of an accommodation (e.g., reading aloud and individual test administration) need to be assessed separately whenever possible. Second, the student populations under study need to be referenced not only with regard to their disability status but also with regard to other relevant variables, for example, the students' level of reading skill, their prior experience with the accommodation, and the degree of improvement in performance that they experienced when using the accommodation. Third, accommodation effects can be investigated using alternatives to group comparison designs, for example, multiple baseline designs across participants.

### Implications for Practice

Overall, the literature to date has shown weak effects for accommodations (Chiu & Pearson, 1999; Tindal, 2002). However, a weak effect for a population of students (e.g.,

students with LD) does little to inform the selection of an accommodation for a particular student. That is, in the absence of any other information, the mean accommodation boost associated with the assignment of an accommodation to a group of students provides our best estimate of the impact for any individual student. However, prediction will improve tremendously if information is available on the students' prior experience with a particular accommodation, especially if the student was afforded multiple assessment opportunities both in the classroom and in test situations using a variety of accommodations (cf. Helwig & Tindal, 2003; Tindal et al., 1998).

A caution in the implementation of test accommodations, especially on high-stakes tests, is that accommodations are not a remedy for low levels of skill on the construct that is being assessed. The large amount of attention being paid to providing students with disabilities with appropriate accommodations may suggest to some students and their families that the "right" combination of accommodations will result in students achieving an adequate level of performance on a test. In this regard, we remind the reader that the passages on our test were two or more grade levels below the level of most of the passages that students would encounter on the statewide assessment. Moreover, as pointed out by Elliott et al. (2002), an accommodation may in fact remove a disability-related barrier for the student tested yet still not have a significant effect on scores.

In conclusion, this study adds another piece to the experimental literature on testing accommodations for students with disabilities. The findings of the study are consonant with previous research in suggesting that the challenge of assigning the most effective and appropriate testing accommodations for students with disabilities, like that of designing the most effective and appropriate instructional programs for these students, is unlikely to be successfully addressed by dictums affecting entire populations of students defined by their category of disability. Instead, much more attention will need to be paid to individual students' characteristics and responses to accommodations in relation to particular types of tests and testing situations.

### ACKNOWLEDGMENTS

This research was supported by Florida Department of Education Grants 874-26210-1RC02 and 874-26220-2R002 to Batya Elbaum.

### REFERENCES

- Campbell, D., & Kenny, D. (1999). *A primer on regression artifacts*. New York: Guilford.
- Chiu, C. W. T., & Pearson, P. D. (1999, June). *Synthesizing the effects of test accommodations for special education and limited English proficient students*. Paper presented at the National Conference on Large Scale Assessment, Snowbird, UT.
- Elliott, S. N., McKeivitt, B. C., & Kettler, R. J. (2002). Testing accommodations research and decision-making: The case of good scores being highly valued but difficult to achieve for all students. *Measurement and Evaluation in Counseling and Development*, 35, 153-166.
- Elliott, S. N., & Roach, A. T. (2002, April). The impact of providing testing accommodations to students with disabilities. In Z. Stevenson (Chair), *How federal requirements are affecting inclusion of special needs stu-*



- dents on state assessments*. Symposium conducted at the meeting of the Annual Convention of the American Educational Research Association, New Orleans, LA.
- Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, 123, 71–99.
- Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgements about test accommodations. *Exceptional Children*, 67, 67–81.
- Goldman, S. R., Hogaboam, T. W., Bell, L. C., & Perfetti, C. A. (1980). Short-term retention of discourse during reading. *Journal of Educational Psychology*, 72, 547–655.
- Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart & Winston.
- Helwig, R., & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children*, 69, 211–225.
- Hollenbeck, K., Rozek-Tedesco, M. A., Tindal, G., & Glasgow, A. (2000). An exploratory study of student-paced versus teacher-paced accommodations for large-scale math tests. *Journal of Special Education Technology*, 15(2), 29–38.
- Individuals with Disabilities Education Act Amendments of 1997, 20 U.S.C. § 1400 *et seq.* (West 1997).
- Marquart, A. M. (2000). *The use of extended time as an accommodation on a standardized mathematics test: An investigation of effects on scores and perceived consequences for students of various skill levels*. Unpublished doctoral dissertation, University of Wisconsin, Madison.
- McKevitt, B. C., & Elliott, S. N. (2003). Effects and perceived consequences of using read-aloud and teacher-recommended testing accommodations on a reading achievement test. *School Psychology Review*, 32, 583–600.
- Morris, R. D., Steubling, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., et al. (1998). Subtypes of reading disability: Variability and a phonological core. *Journal of Educational Psychology*, 80, 347–373.
- No Child Left Behind Act of 2001. Retrieved July 22, 2004, from <http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf>
- Phillips, S. E. (1994). Testing condition accommodations: Validity versus disabled rights. *Applied Measurement in Education*, 7, 93–120.
- Thurlow, M. L., & Bolt, S. (2001). *Empirical support for accommodations most often allowed in state policy* (Synthesis Rep. No. 41). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved January 15, 2003, from <http://www.coled.umn.edu/NCEO/OnlinePubs/Synthesis41.html>
- Thurlow, M. L., House, A. L., Scott, D. L., & Ysseldyke, J. E. (2000). Students with disabilities in large-scale assessments: State participation and accommodation policies. *Journal of Special Education*, 34, 154–163.
- Thurlow, M. L., McGrew, K. S., Tindal, G., Thompson, S. J., Ysseldyke, J. E., & Elliott, J. L. (2000). *Assessment accommodations research: Considerations for design and analysis* (Tech. Rep. No. 26). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Tindal, G. (2002). How will assessments accommodate students with disabilities? In R. W. Lissitz & W. D. Schafer (Eds.), *Assessment in educational reform: Both means and ends* (pp. 100–123). Boston: Allyn & Bacon.
- Tindal, G., & Fuchs, L. (1999). *A summary of research on test changes: An empirical basis for defining accommodations*. Lexington, KY: Mid-South Regional Resource Center.
- Tindal, G., Heath, B., Hollenbeck, K., Almond, P., & Harniss, M. (1998). Accommodating students with disabilities on large-scale tests: An experimental study. *Exceptional Children*, 64, 439–450.

Copyright of Exceptionality is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.