The Journal of Special Education

http://sed.sagepub.com

Effects of an Oral Testing Accommodation on the Mathematics Performance of Secondary Students With and Without Learning Disabilities

Batya Elbaum

J Spec Educ 2007; 40; 218
DOI: 10.1177/00224669070400040301

The online version of this article can be found at:
http://sed.sagepub.com/cgi/content/abstract/40/4/218

Published by:
Hammill Institute on Disabilities

and

http://www.sagepublications.com

Effects of an Oral Testing Accommodation on the Mathematics Performance of Secondary Students With and Without Learning Disabilities

Batya Elbaum
University of Miami

This study compared the performance of students with and without learning disabilities (LD) on a mathematics test using a standard administration procedure and a read-aloud accommodation. Analyses were conducted on the test scores of 625 middle and high school students (n = 388 with LD) on two equivalent 30-item multiple-choice tests. Whereas mean scores for students both with and without LD were higher in the accommodated condition, students without disabilities benefited significantly more from the accommodation (ES = 0.44) than students with LD (ES = 0.20). In addition, effect sizes from the present study were combined meta-analytically with those of previous studies. Results of the meta-analysis revealed that for elementary students, oral accommodations on a mathematics test yielded greater gains for students with LD than for students without disabilities; for secondary students, the converse was true. Findings of the study are discussed in relation to the question of the validity of an oral accommodation on mathematics tests for students both with and without disabilities.

One of the most important accomplishments of recent U.S. federal education legislation has been to promote the full participation of students with disabilities in state educational accountability systems. The Individuals with Disabilities Education Act (IDEA) Amendments of 1990, the No Child Left Behind Act of 2001, and the most recent reauthorization of IDEA as the Individuals with Disabilities Education Improvement Act (IDEIA; 2004) have affirmed the principle of including students with disabilities in statewide assessments, as well as the need to offer appropriate accommodations or alternate testing procedures, as necessary, to support students' participation.

Over the past 15 years, the National Center on Educational Outcomes (NCEO) has been documenting states' policies and practices regarding the participation of students with disabilities on statewide assessments. In its latest report, Thompson, Johnstone, Thurlow, and Altman (2005) stated that one of the six key factors cited by states as contributing to positive trends in the participation and performance of students with disabilities has been the development and provision of accommodation guidelines and training. In addition, recent studies have catalogued the rapidly evolving use of accommodations on statewide tests (e.g., Johnson, Kimball, Brown, & Anderson, 2001; Thurlow, House, Scott, & Ysseldyke, 2000).

Given the serious consequences of test outcomes for states, districts, schools, and individual students, the validity of interpretations of test scores when students are given particular accommodations has been a critical question in both the research and policy arenas (Thurlow & Bolt, 2001; Thurlow, House, et al., 2000; Thurlow, McGrew, Tindal, Thompson, Ysseldyke, & Elliott, 2000; Tindal, 2002; Tindal & Fuchs, 1999). There is general consensus that to be considered a valid accommodation, a modification in test administration should remove disability-related variance without affecting construct-relevant variance. For example, allowing students with motor difficulties to dictate their solutions to mathematics problems to a scribe addresses the students' specific disability without affecting their mathematics skills. This accommodation would be expected to improve the test performance of students with motor impairments only. If the accommodation were given to students without motor impairments, no impact on test performance would be expected to result.

One test of the validity of a testing accommodation is whether it changes the meaning of test scores as evidenced by variance in factor structure or differential item functioning across tests administered with and without accommodations. Pomplun and Omar (2000) investigated the factorial structure of a fourth-grade state mathematics assessment administered to three groups of students: general education students taking the test without accommodations, students with LD taking the test without accommodations, and students with LD taking the test with a read-aloud accommodation. Results indicated

Address: Batya Elbaum, Department of Teaching and Learning, School of Education Research, University of Miami, 1551 Brescia Ave., Coral Gables, Florida 33146. E-mail: elbaum@miami.edu

the invariance of the test's factor structure across all three groups, providing support for the comparability of scores under both testing conditions and the validity of aggregating the scores of students with and without disabilities. A similar finding of invariance in factor structure was reported by Huynh, Meyer, and Gallant (2004) for an eighth-grade mathematics test administered to general education students and also to students with disabilities, with and without an oral accommodation.

Fuchs (2000) examined whether various testing accommodations were associated with differential item functioning (DIF) for students with disabilities compared to students without disabilities. In general, item functioning should be invariant in regard to characteristics of test-takers that are not related to the construct being measured, for example, gender and ethnicity. However, item functioning is expected when an accommodation removes construct-irrelevant variance. Thus, if reading ability is irrelevant to a measure of mathematics ability, then a read-aloud accommodation should change item functioning only for those test-takers with poor reading ability. Results showed that 50% of concepts and applications items showed evidence of DIF such that students with LD had improved performance on those items with a read-aloud accommodation.

The most widely discussed way of evaluating the validity of a testing accommodation for students with disabilities is by examination of the disability status by testing condition interaction effect. The strong form of the interaction hypothesis is that a valid accommodation should improve the performance of students with disabilities while having no effect on the performance of students without disabilities. Based on their review of the research, Sireci, Scarpati, and Li (2005) concluded that the "interaction hypothesis needs qualification" (p. 481). Consistent with the concept of "differential boost" (Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000), the hypothesis advocated by Sireci et al. is that a valid accommodation should improve the performance of students with disabilities to a significantly greater extent than it improves the performance of students without disabilities:

> When [students with disabilities] . . . exhibit greater gains with accommodations than do their general education peers, an interaction is present. When gains experienced by [students with disabilities] are significantly greater than the gains experienced by their general education peers, the fact that the general education students achieved higher scores with an accommodation condition does not imply that the accommodation is unfair. It could imply that the standardized test conditions are too stringent for *all* students. (p. 481)

Results of the research on oral accommodations for mathematics tests are by no means unequivocal. Some studies, but far from all, have reported a significant positive effect of the accommodation for students with disabilities, with little or no effect for students without disabilities. For example, Tindal, Heath, Hollenbeck, Almond, and Harniss (1998) studied the mathematics test performance of fourth-grade students with and without disabilities. Students with IEPs (Individualized Education Programs) in reading or math achieved significantly higher scores when the test was read aloud to them than when they read the test items themselves. The accommodation effect size for students with disabilities was 0.82, compared to –0.18 for students without disabilities.

Other students have found significant positive effects for students with and without disabilities, as well as a significant interaction effect. For example, Weston (2002) analyzed the performance of fourth-grade students with and without LD on two parallel forms of the mathematics portion of the National Assessment of Educational Progress. Both groups showed gains with an oral accommodation, and students with disabilities gained significantly more than did students without disabilities, $ES = 0.64$ versus $ES = 0.31$, respectively. Weston also examined the relationship between students' reading level and the gain experienced as a result of the accommodation. Students with higher reading proficiency gained less than did students with lower proficiency. Additionally, the read-aloud accommodation improved performance on word problems more than it did on calculation problems.

Still other studies have found a significant positive effect for all students, with no significant difference in the magnitude of the effect for students with and without disabilities. For example, Meloy, Deville, and Frisbie (2000) investigated the effect of a read-aloud accommodation on the performance of middle school students on the *Iowa Tests of Basic Skills,* which included a test of Math Problem-Solving and Data Interpretation. Students with a reading disability showed a gain of 0.75 *SD*, while the gain for students without disabilities was approximately 0.50 *SD*. The interaction effect was not statistically significant. Similarly, Johnson (2000), in a study of the impact of an oral accommodation on the performance of fourth-grade students with and without reading disabilities on the *Washington Assessment of Student Learning* math test, also failed to find a significant interaction of disability status and testing condition.

Several studies have found no significant accommodation effect for either group of students, or an effect in the unanticipated direction. For example, Helwig and Tindal (2003) studied elementary and middle school students' math test performance in a standard and a read-aloud condition. Contrary to expectations, students with low reading skills, but adequate math skills, performed better in the standard condition than in the read-aloud condition.

Finally, a small number of studies have reported mixed results depending on item characteristics. The hypothesis under consideration is that the magnitude of the accommodation boost depends on the reading difficulty of test items, such that the effect of a read-aloud accommodation would be more marked for items with greater linguistic complexity. Helwig, Rozek-Tedesco, Tindal, Heath, and Almond (1999) investigated the impact of a video-presented, read-aloud accommodation on a mathematics test administered to 325 sixth graders, 12% of whom were students receiving special education services. Overall, the investigators did not find a statistically significant difference in students' performance across conditions. In particular, no accommodation effect was found for low reading-fluency students when performance on all 60 test items was examined. Students' performance was also examined only on the 6 items identified as posing significant linguistic challenges, defined as involving sentence length of greater than 40 words, five or more verbs in the passage, and at least three words familiar to fewer than 90% of sixth-grade students. On these items, low-fluency readers with average or above-average math skills profited significantly from the read-aloud accommodation. In another example, Schulte, Elliott, and Kratochwill (2001) examined the effects of testing accommodations on standardized mathematics test scores of fourth graders with and without disabilities. On multiple-choice items, students with disabilities benefited more from a read-aloud administration than did students without disabilities (*ES* = 0.41 vs. *ES* = 0, respectively). In contrast, for the constructed response items, effect sizes were similar for both groups (*ES* = 0.31 vs. *ES* = 0.35 for students with and without disabilities, respectively).

Given the diverse findings of previous research, and the relative dearth of studies targeting older students, the present study was designed to address the following research questions: What is the effect of an oral testing accommodation on the performance of middle and high school students with and without LD on a mathematics test? As a group, do students with LD benefit more from the accommodation than students without disabilities? Applying an individual difference perspective, what percentage of students in each group experience an improvement, and what percentage experience a decrement in performance, as a result of the accommodation? When the results of this study are added to those of previous research, what general conclusions can be drawn concerning the validity of an oral accommodation for students with and without disabilities?

Method

Determination of Sample Size

The minimum number of participants to be included in this study was determined based on a power analysis (Cohen, 1988) that took into account both the type of statistical test that would be performed (a repeated measures analysis of variance [ANOVA]) and the magnitude of the accommodation effect sizes found in previous research (ranging from approximately −0.20 to 0.80). For example, given a significance level of p = .05, single-tailed, an analysis conducted with 100 students in each group would have a probability of .41 of detecting an *ES* of 0.20, a probability of .88 of detecting an *ES* of 0.40, and a probability of over .99 of detecting an *ES* greater than 0.60. If the true effect size for the accommodation were small, a larger sample would be needed to ensure detection of a significant effect. With n = 300 in each group, the respective probabilities rise to .79, > .99, and > .99 for effect sizes of 0.20, 0.40, and 0.60, respectively. Therefore, the study was designed so as to obtain data from approximately 600 students.

Participants

Participants were 643 students in Grades 6 through 10 (58% boys, n = 391 identified as students with LD) attending three public middle schools and three public high schools in a large, metropolitan school district in the southeastern United States. The six schools from which participants were recruited had student populations ranging in size from 1,207 to 4,655. The percentage of students receiving free or reduced-price lunch ranged from 33% to 89% across the six schools. White, non-Hispanic students made up from 3% to 15% of the population at each school; Black, 2% to 38%; and Hispanic, 45% to 95%. Limited English proficient students made up between 5% and 22% of students at each school. Table 1 presents the distribution of students by grade and disability status.

The school district followed federal guidelines for the identification of students with LD, requiring evidence of a discrepancy between the student's level of intellectual function-

TABLE 1. Distribution of Participants by Grade and Disability Status

	Grade			Total	Grade		Total	
	6	7	8	middle school	9	10	high school	Total
LD[a]	45	95	47	187	171	33	204	391
SWOD[b]	18	85	37	140	103	9	112	252
Total	63	180	84	327	274	42	316	643

[a]LD = students with learning disabilities. [b]SWOD = students without disabilities.

ing and achievement on tasks required for basic reading skills, reading comprehension, oral expression, listening comprehension, mathematics calculation, mathematics reasoning, or written expression. The magnitude of the discrepancy required for eligibility differed by age, with the minimum required discrepancy defined as "significant" for children under the age of 7, 1 *SD* for students ages 7 to 10, and 1.5 *SD* for students ages 11 and above.

Individual measures of students' prior mathematics and reading achievement were not available. However, a general framework for interpreting the performance of students who participated in this study can be obtained from state-reported data, disaggregated by grade level within each school, on student performance on the state's annual reading and mathematics assessments. The state defines scores at or above Level 3 in the state's five-level system as demonstrating adequate proficiency, and requires students to demonstrate such proficiency on the 10th-grade test to graduate with a standard diploma. At the six schools from which participants were drawn for this study, the percentage of students who scored at Level 3 or above varied somewhat by grade and school. In reading, the percentages of students without disabilities who demonstrated adequate proficiency ranged from 31% to 57% for Grades 6, 7, and 8 and from 16% to 38% for Grades 9 and 10. The corresponding percentages for students with LD were 3% to 10% in Grades 6, 7, and 8 and 1% to 5% in grades 9 and 10. On the mathematics assessment, the percentages of students without disabilities scoring at Level 3 or above were 21% to 59% in Grades 6, 7, and 8, and 34% to 65% in Grades 9 and 10. For students with LD, the percentages were 2% to 15% in Grades 6, 7, and 8, and 5% to 22% in Grades 9 and 10.

The state also administers an annual norm-referenced assessment in reading and mathematics. The norm-referenced assessment administered in the year this study was conducted was the *Stanford Achievement Test* (9th ed.; SAT-9; 1996). The state reports the median national percentile rank (NPR) for each grade level within a school. At the six schools involved in this study, the median NPRs in reading ranged from 31 to 59 in Grades 6, 7, and 8 and from 30 to 37 in Grades 9 and 10. The NPRs in mathematics ranged from 40 to 67 in Grades 6, 7, and 8 and from 46 to 60 in Grades 9 and 10.

Mathematics Assessment Instrument

The mathematics tests used in this study were constructed so as to meet several criteria. First, given the relevance of this study to decisions about accommodations for students with disabilities on statewide tests, the assessments needed to be broadly similar in content, presentation format, and response format to the multiple-choice sections of the statewide mathematics assessment. Second, given the fully counterbalanced design of the study, two alternate forms of equivalent difficulty were needed. Third, the difficulty level of the assessment needed to be targeted to the skill level of the participating students, in that it would be impossible to assess the impact of an oral accommodation on students whose test performance without any accommodations was already at the top of the scale.

Sixty test items were drawn from various practice materials for the statewide test that were not in use at any of the schools where the study was conducted. The items addressed fifth-grade state standards in the domains of number sense and operations, geometry, data analysis and probability, algebraic thinking, and measurement. Each item consisted of a problem statement and four answer choices. Some items included diagrams (28%), tables (10%), or graphs (10%).

The 60 test items were piloted with groups of elementary, middle, and high school students with LD who would not be participating in the study. The percentage of students responding correctly to an item was used as an index of item difficulty. The items were ordered by difficulty and then alternately assigned to one of two test forms. Slight modifications in item assignment were made to make the forms as equivalent as possible with respect to item type (multiplying fractions, reading graphs, etc.) and linguistic difficulty (see below). On both test forms, the items were ordered from least to most difficult.

Linguistic Difficulty. Linguistic difficulty of the items was assessed according to the procedure used by Helwig et al. (1999). Each of the 60 items was analyzed for the total number of words, verb phrases, and difficult words present. The total number of words for each item was determined by counting each word in the problem statement and all answer choices. Proper nouns, abbreviations, and numbers appearing in Arabic form were not counted. In addition, if a word or phrase appeared three or more times in an item, it was counted only the first time. Difficult words were identified using *The Living Word Dictionary* (Dale & O'Rourke, 1979), which contains more than 43,000 words. Each entry provides a grade level at which a percentage of students correctly identified the definition of the word. Difficult words were designated as those familiar to less than 90% of sixth-grade students.

Items included in Forms A and B had an average of 18 and 19 words per item, respectively, ranging from 5 to 35 words for Form A and from 4 to 44 words for Form B. In the Helwig et al. (1999) accommodation study, criteria for a linguistically challenging math item for middle-school students were met when the item contained at least 34 words, four verbs, and one difficult word. In the present study, this criterion was not met for any items on Math Form A (though four items met two out of the three criteria), and was met once on Form B (four other items met two out of the three criteria). Thus, by this measure, the items utilized in the present study presented less reading difficulty than did those administered by Helwig et al. (1999).

Verification of Form Equivalence. The pilot test forms were administered under a standard testing condition to several samples of students at schools that were demographically comparable to those participating in the study. For a sample of 34 middle-school, general education students, the correlation between scores on the two forms of the test was $r = .80$, $p < .001$. For a sample of 31 high school students with LD, the correlation was $r = .77$, $p < .001$. Additionally, a separate sample of high school, general education students ($n = 53$) and students with LD ($n = 70$) was randomly assigned to either Form A or to Form B. Two separate *t*-tests for independent samples—one for students with LD and one for general education students—indicated no statistically significant difference between students' mathematics performance on the two forms.

Procedure

Recruitment. Approval to conduct the study was obtained from both the university Institutional Review Board and the local school district. Permission to recruit students for the study was obtained from principals of the six participating schools. Individual teachers within each school had the option of allowing their classes to participate.

In the schools where this study was conducted, some students with LD received mathematics instruction in general education classrooms, whereas others received mathematics instruction in a special education class. Thus, recruitment efforts were conducted in both general and special education mathematics classrooms. However, only special education classes that followed the regular mathematics curriculum were targeted. Additionally, because the sampling plan called for approximately equal numbers of students with and without LD, the number of students with LD recruited for the study represented a large proportion, varying from 31% to 55%, of the students with LD at each school.

A member of the research team visited each participating mathematics classroom to give a brief explanation of the study and to distribute parental consent forms. Students who returned signed forms indicating parental consent to participate were asked to sign an informed assent form.

Teachers were given a block of code numbers and asked to assign a number to each participating student and to write the number on his or her test booklet. On a separate sheet of paper listing the code numbers for participating students, the teacher indicated whether the student had an IEP, and if so, the category of disability under which the student was deemed eligible for special education services. This procedure made it possible to correctly identify target students without retaining any individually identifying information in study records. A small number of students, subsequently identified as students with disabilities other than LD, took the tests; however, their tests were not scored and their data were not used in the study.

Test Administration. The two mathematics assessments were group-administered to students in their general education mathematics classes during a single class period. In each classroom, the teacher was asked to give each student a test booklet, as well as an answer sheet bearing his or her identification number. Separate test booklets and answer sheets were used for each test. The first page of the test booklet had a sample item and accompanying response choices; subsequent pages had two test items per page. The test administrator read a scripted introduction and then reviewed the sample problem.

For the standard condition, students were instructed to go through the problems as they would on any test, working out as many problems as they could in the total allowed time of 25 min. In untimed field testing, this time was found to be sufficient to allow all students to complete the 30-item test.

For the read-aloud condition, students were encouraged to follow along with the test administrator, who read each item aloud twice. After each item was read, the administrator gave students a set amount of time—45, 60, or 75 s, depending on item difficulty—to complete the problem. In untimed field testing, these times were found to be sufficient to allow all students to complete each item. Students were instructed not to turn to the next page until told to do so.

In accordance with the procedure used for the statewide mathematics assessment, students were not permitted to use calculators.

Counterbalancing of Forms and Order of Test Conditions. Classrooms were randomly assigned to one of four different form-by-testing condition combinations: Form A/ standard administration, followed by Form B/oral accommodation; Form A/oral, Form B/standard; Form B/standard, Form A/oral; Form B/oral, Form A/standard. This procedure resulted in a relatively equal distribution of students with and without LD across combinations (students with LD: 80, 101, 115, and 77 in the four conditions; students without disabilities: 64, 60, 51, and 61).

Results

The first step in the analysis was to examine whether there were any differences in students' test performance owing to form or test condition order. Two separate one-way ANOVAs, one for students with LD and one for students without disabilities, indicated that there were no statistically reliable differences in students' performance owing to form-by-order condition. For students with LD, $F(3, 369) = 0.35$, $p = .79$; for students without disabilities, $F(3, 232) = 0.45$, $p = .72$.

The second step was to ensure that the analysis of accommodation effects included only those students for whom it would be possible to measure an improvement in performance on the test. For students whose scores in the unaccommodated condition were already at or near the maximum score possible, improvements owing to the read-aloud accommodation would be subject to a ceiling effect. A total of 18 students, including 3 with LD, had scores of 25 or above in the standard condition. These 18 students, representing 2.7% of study participants, were not included in subsequent analyses.

Scores for the 625 students remaining in the analysis are presented in Table 2. Two independent-sample *t* tests were

TABLE 2 . Means and Standard Deviations of Students' Test Scores in Standard and Read-Aloud Conditions

	Students with LD				Students without disabilities			
	Standard	Read-Aloud	Boost	Effect size	Standard	Read-Aloud	Boost	Effect size
Middle school	10.93 (4.19)	11.59 (4.36)	0.66 (3.82)	0.16	16.70 (4.65)	18.84 (4.31)	2.13 (4.08)	0.46
High school	13.40 (4.37)	14.47 (4.30)	1.07 (3.75)	0.24	17.96 (4.46)	19.87 (4.73)	1.91 (3.49)	0.43
Total	12.21 (4.45)	13.08 (4.56)	0.87 (3.78)	0.20	17.25 (4.60)	19.29 (4.52)	2.04 (3.83)	0.44

Note. Middle school students with LD: n = 187; middle-school students without disabilities: n = 134; high school students with LD: n = 201; high school students without disabilities: n = 103.

conducted to investigate whether the accommodation boost differed for middle- and high school students. There were no statistically significant differences for either students with LD, $t(386) = -1.07$, $p = .28$, or students without disabilities, $t(235) = 0.44$, $p = .66$. Therefore, students' school level was not entered as a variable in subsequent analyses.

Accommodation effect sizes were calculated as the accommodation boost divided by the standard deviation in the unaccommodated condition. The mean accommodation boost was 0.87 points, $ES = 0.20$, for students with LD, compared to 2.04 points, $ES = 0.44$, for students without disabilities.

A 2 × 2 repeated measures ANOVA with one between-subjects factor (disability status: LD vs. no disabilities) and one within-subjects factor (test condition: standard vs. read-aloud) revealed the expected statistically significant main effect for disability status, $F(1623) = 275.56$, $p < .001$, partial $\eta^2 = .31$. Partial η^2 is the proportion of the effect plus error variance that is due to the specified effect. In addition, there was a statistically significant main effect for test condition, $F(1623) = 86.21$, $p < .001$, partial $\eta^2 = .12$. The disability by test condition interaction was also statistically significant, $F(1623) = 13.87$, $p = < .001$, partial $\eta^2 = .02$. Overall, students without disabilities benefited more from the read-aloud accommodation than did students with LD. The differential effect of the oral accommodation, as indexed by η^2, accounted for a very small proportion of the overall variance in mathematics performance.

To investigate the relationship between the impact of the oral accommodation and students' mathematics skills, correlations were computed separately for students with and without LD between the accommodation boost students received and their scores in the accommodated condition. The resulting correlations were $r = .44$, $p < .001$, and $r = .40$, $p < .001$, for students with and without LD, respectively. Additionally, accommodation effect sizes were calculated separately for students with and without LD performing above, or below, the 50th percentile on the test in the accommodated condition. Students with LD in the top half of the score distribution ($n = 115$) showed a mean accommodation effect size of 0.61, compared to $ES = 0.02$ for students with LD ($n = 273$) in the lower half of the distribution. For students without disabilities, the corresponding effect sizes were 0.55 for students in the upper half of the distribution ($n = 189$), compared to $ES = 0.11$ for students in the lower half of the distribution ($n = 48$). These results suggest that regardless of disability status, the stronger the students' mathematics skills, the more substantial the benefit they derive from a read-aloud accommodation.

Given that mean measures of the accommodation effect do not capture the heterogeneity of students' responses to the accommodation, an additional analysis was conducted in which students were categorized as having experienced a benefit to their test performance as a result of the accommodation, a detriment to their performance, or no difference. For the purpose of this analysis, the relatively stringent criterion of a 1 *SD* shift was applied, in that a shift of this magnitude, corresponding to a move from the 50th to the 84th percentile on a test, or, conversely, from the 50th percentile to the 16th, represents a change substantial enough to affect such real-life consequences as grade promotion or graduation with a standard diploma. In a similar analysis, Helwig and Tindal (2003)

TABLE 3. Cross-Tabulation of Accommodation Impact for Students With LD and Students Without Disabilities

	Impact of accommodation			
Disability status	Benefit	No difference	Detriment	Total
Students with LD				
Count	60	298	30	388
% within disability	15.5	76.8	7.7	100.0
% within impact	51.7	63.9	69.8	62.1
% of total	9.6	47.7	4.8	62.1
Students without disabilities				
Count	56	168	13	237
% within disability	23.6	70.9	5.5	100.0
% within impact	48.3	36.1	30.2	37.9
% of total	9.0	26.9	2.1	37.9
Total				
Count	116	466	43	625
% within disability	18.6	74.6	6.9	100.0
% within impact	100.0	100.0	100.0	100.0
% of total	18.6	74.6	6.9	100.0

used the less stringent criterion of .5 *SD* shift in test score. As seen in Table 3, whereas 15.5% of students with LD experienced a benefit from the accommodation, 7.7% experienced a detriment to performance. Of students without disabilities, 23.6% experienced a benefit, whereas 5.5% showed a detriment. The difference in the distribution of impact was statistically significant, $\chi^2(2) = 7.06$, $p = .03$.

To address the final research question, accommodation effect sizes were derived from all available studies of read-aloud accommodations on mathematics tests for students with disabilities or poor readers. The results, grouped by students' school level (elementary vs. secondary), are displayed in Table 4. Details regarding the specific data and formulas used for computation of effect sizes are available from the author. As seen in Table 4, six studies involved only elementary students, five studies involved only secondary students, and three studies included both elementary and middle school students. The accommodation effect sizes for elementary school students with disabilities ranged from 0.10 to 0.82, $M = 0.37$, $SD = 0.26$. For secondary students with disabilities/low readers, accommodation effect sizes ranged from −0.07 to 0.30, $M = 0.10$, $SD = 0.12$. Within-study differences in the accommodation effect size for students with disabilities/low readers and students without disabilities/high readers ranged from 0 to 1.0 for elementary school students and −0.27 to 0.02 for secondary students. Thus, effect size differences for groups of elementary students were either nil or favored students with disabilities; differences for secondary students were either nil or favored students without disabilities.

The study-level effect size differences were submitted to a meta-analysis following procedures described in Cooper (1998). Each accommodation effect size difference was first multiplied by the inverse of its variance. This procedure results in greater weight being accorded to larger samples, which yield more stable estimates of population parameters. The mean weighted effect size difference, d, across all studies was $d = 0.03$. For studies of elementary students ($k = 8$), $d = 0.20$, with a 95% confidence interval of +/− .10. For studies of secondary students ($k = 6$), $d = -0.12$, with a 95% confidence interval of +/− .09. The fact that the aforementioned confidence intervals do not include zero indicates that both effect size differences were reliably different from zero. The homogeneity statistic, Q, which is evaluated as χ^2 with k−1 degrees of freedom, was statistically significant, $Q(13) = 55.78$, $p < .01$, indicating the presence of greater variation than would be expected based on chance alone. When school level was examined as a moderator variable, the combined Q statistics for the elementary and secondary groups, $Q(7) = 28.22$, and $Q(5) = 5.63$, subtracted from the Q value for all effect sizes combined, was $Q = 55.78 - 33.85 = 21.93$. This value, evaluated as a chi square with $df = 1$, was statistically significant at $p < .01$, indicating a statistically significant association of students' school level with the difference in effect sizes for students with and without LD.

Discussion

The purpose of this study was to examine the effect of an oral testing accommodation on the performance of students with and without LD on a mathematics test similar in content and format to portions of typical statewide mathematics assessments. This study implemented several recommendations in the literature (see Helwig & Tindal, 2003; Sireci et al., 2005). First, the study targeted middle- and high school students, whose performance under accommodated testing conditions has not been as frequently investigated as that of elementary students. Second, the study was carefully targeted to the students' mathematics ability, so as to provide the most favorable conditions possible for detection of an accommodation effect. Third, the study employed a repeated measures design such that each student served as his or her own control.

Findings of this study showed that a read-aloud accommodation on a mathematics test resulted in improved performance for students both with and without disabilities. Students without disabilities profited more from the accommodation, on average, than did students with LD. Whereas students with LD showed an average gain of approximately .2 *SD*, students without disabilities showed a gain of twice that magnitude, or about .4 *SD*. Over 23% of general education students realized an improvement of 1 *SD* or more, whereas only 15% of students with LD derived a benefit of this same magnitude.

The observed interaction effect, operating in the unanticipated direction, suggests that the accommodation did not address a specific disability-related characteristic of the students with LD. Indeed, the statistically significant difference in performance in favor of students *without* disabilities suggests that the accommodation removed some impediment to performance that was shared by some students in both groups. To the extent that the oral accommodation addressed a barrier resulting from poor reading ability, the implication is that the general education students in this study included a large number of poor readers. Although individual measures of students' reading proficiency were not obtained as part of this study, the statewide assessment data reported earlier support the notion that a large number of the general education students who took part in this study likely have inadequate reading skills. On average, only one-third to one-half of middle-school students without disabilities, and one-sixth to one-third of the high school students without disabilities, passed the state's reading assessment. These students, although not formally identified as having a reading disability, would clearly be defined as poor or very poor readers by the state's yardstick.

The results of the correlational analysis offer a complementary explanation of the overall findings. Students with stronger mathematics skills, whether they were students with an identified LD, benefited more from the oral accommodation. This finding suggests that the significant interaction effect, whereby general education students benefited more from

TABLE 4. Studies of Read-Aloud Accommodations on Mathematics Tests for Students With Disabilities/Low Readers

Study	Target group	Target group *n*	Comparison group	Comparison group *n*	Target group effect size	Comparison group effect size	Effect size difference	Statistically significant interaction effect?
Elementary students								
Fuchs et al. (2000)	LD	192	SWOD	181	.36	.36	.00	No
Helwig & Tindal (2003)	SWD	104	SWOD	533	.13	.07	.06	NR
Helwig et al. (2002)	LD-R	88	SWOD	88	.17	−.08	.25	NR
Johnson (2000)	RD	38	SWOD	38	.56	−.08	.64	No
Pomplun & Omar, 2000	LD	1,542[a]			.10			NA
Schulte et al. (2001)	SWD	43	SWOD	43	.40	.21	.19	No
Tindal (2002)	SWD	104	SWOD	575	.15	.02	.13	Yes
Tindal et al. (1998)	SWD	42[b]	SWOD	122[c]	.82	−.18	1.00	Yes
Weston. (2002)	LD	65	SWOD	54	.64	.31	.33	Yes
Secondary students								
Calhoun et al. (2000)	LD-R	81			.30			NA
Elbaum (2006)	LD	388	SWOD	237	.20	.44	−.24	Yes
Helwig & Tindal (2003)	SWD	141	SWOD	440	−.07	−.03	−.04	NR
Helwig et al. (1999)	Low readers	94	Avg-high readers	45	.06	.04	.02	NR
Helwig et al. (2002)	LD	102	SWOD	102	−.03	.00	−.03	NR
Hunyh et al. (2004)	SWD	3,576[d]			.21[e]			NA
Meloy et al. (2000)	LD-R	62[f]	SWOD	198[g]	.08	.35	−.27	No
Tindal (2002)	SWD	111	SWOD	513	.08	.11	−.03	No

Note. SWD = students with disabilities; SWOD = students without disabilities; LD = students with learning disabilities; LD-R = students with a learning disability in reading; RD = students with a reading disability; NR = not reported; NA = not applicable.

[a]Standard administration, $n = 1{,}369$; oral administration, $n = 173$. [b]Standard administration, $n = 22$; oral administration, $n = 20$. [c]Standard administration, $n = 89$; oral administration, $n = 33$. [d]Standard administration, $n = 2{,}642$; oral administration, $n = 934$. [e]Adjusted ES from ANCOVA using students' score on a reading test as the covariate. [f]Standard administration, $n = 29$; oral administration, $n = 33$. [g]Standard administration, $n = 98$; oral administration, $n = 100$.

the accommodation than did students with LD, may be due to the greater proficiency in mathematics of students without disabilities. This finding would lend further support to the idea that providing access to test items by removing reading ability as a barrier will differentially improve the performance of students who have a higher level of skill in the content area being tested.

Another contributing explanation may be that the read-aloud condition enhanced attention to the problem statements and response choices, resulting in fewer errors owing to failure to encode the task correctly or to make careful distinctions among response choices. Support for this idea comes from the observation that the accommodation had an overall positive effect on performance even though, by criteria described in the previous literature (Helwig et al., 1999), the items did not present substantial reading challenges. Weston (2002) also found that a read-aloud accommodation had a positive effect on performance on calculation-only items, which do not require processing of text. Weston offered the explanation that the equivalent gain for students with and without LD had to do with the fact that "students are kept on task when the test is read aloud" (p. 19). This would still not fully explain, however, the finding of greater benefit of an oral accommodation for students without disabilities.

Validity of the Oral Accommodation

As mentioned earlier, one perspective on the validity of testing accommodation is that in order for the meaning of scores achieved by students with disabilities in an accommodated condition to be the same as that of scores achieved by students in a standard condition, the accommodation should not produce a benefit for students without disabilities (Phillips, 1994). By this criterion, the oral accommodation afforded to students in this study would not be considered valid. Indeed, to the extent that the results of this study are replicated, it may turn out that instead of "leveling the playing field," an oral accommodation on a mathematics test may *increase* the gap in performance between secondary students with and without disabilities. This has clearly not been the intention of federal legislation aimed at promoting the participation of students with disabilities in states' accountability systems.

At this juncture, the argument made by Sireci et al. (2005) warrants serious consideration. If the primary objective of implementing testing accommodations is not that of "closing the gap," but rather that of achieving a more precise measurement of students' abilities in an area such as mathematics, then it would follow that the oral accommodation ought to be regularly offered to all students.

With regard to individual differences, the findings of this study, similar to those of Helwig and Tindal (2003), Elbaum, Arguelles, Campbell, and Saleh (2004), among others, suggest that accommodations are not uniformly benign, but instead have the potential to interfere with performance for some students. Whereas a read-aloud accommodation may remove construct-irrelevant variance owing to reading difficulty, it may also introduce construct-irrelevant variance owing to other factors. For example, teachers whose students participated in the study by Weston (2002) perceived that general education students became impatient with the time needed to finish reading the items aloud and reported that they disliked the pacing of the test in the accommodated condition.

Finally, the results of this study were combined meta-analytically with those of 13 previously published empirical studies. The results of the meta-analysis provide support for the hypothesis that the impact of oral accommodations on students' mathematics performance is not the same for elementary and secondary students. Whereas the accommodation boost for elementary students is clearly of greater magnitude for students with LD than it is for students without LD, the impact on secondary students shows greater benefits for students without disabilities. Though explanations for this finding are unclear, the pattern is consistent enough to warrant further investigation.

Limitations

A limitation of the present study, similar to that of other studies of oral testing accommodations, is the confounding of the accommodation meant to remove reading ability as a factor in performance with concomitant factors that are unrelated to reading ability, such as the pacing and attention-focusing that come with having the test read aloud.

An additional limitation of the study is the fact that no individual-level measure of reading ability was included in the design. Thus, it was not possible to test whether the variation in accommodation effects was associated with variation in individual students' reading skills. At the group level, the reported state assessment data do support the idea that students with LD were poorer readers than their peers without identified disabilities, as one would expect. These data also suggest that there is greater overlap in the distribution of reading proficiency levels across students with and without disabilities than is often acknowledged. If 95% of students with LD were failing the high-stakes high school reading assessment, so were two-thirds of the students without an identified disability.

Implications for Research

Several implications for future accommodations research can be drawn from the present study. First, as noted by previous researchers (e.g., Elliott, Kratochwill, & McKevitt, 2001), the effects of different components of an accommodation, such as oral presentation and pacing, should be independently assessed whenever possible. Second, care should be taken to ensure that the test material is appropriately targeted to all groups of students that are included in the design. Third, factors believed to contribute construct-irrelevant variance—particularly reading ability, insofar as students with LD are concerned—should be measured for each student and included as variables in the analysis. Fourth, the level of linguistic challenge presented by the items should be reported, and controlled across alternate forms. Finally, future research should examine changes in students' response to an accommodation over time, to ascertain whether experience with an accommodation alters the degree to which the student benefits from the accommodation.

Implications for Practice

The findings of this study underscore the need to reframe the issue of testing accommodations as one that is relevant to all students, and not only to students with disabilities. If particular testing accommodations, such as oral presentation of items on a mathematics test, are deemed appropriate for general education students, as well as for students with disabilities, then the incorporation of accommodations into all testing procedures must be conceptualized as part of the universal design of assessment systems (see Thompson, Johnstone, & Thurlow, 2002), rather than as an add-on for special populations.

From a disability perspective, the emphasis should be on the appropriate and individualized assignment of accommodations to students through an informed decision of the IEP team (e.g., Edgemon, Jablonski, & Lloyd, 2006). With regard to mathematics testing in particular, it has been argued that when students are provided a read-aloud testing accommodation, their test performance represents a relatively accurate measure of their mathematics skill (see Weston, 2002). However, given the findings of the present study, it would be inappropriate to conclude that we should simply assign all students with LD to a read-aloud testing condition. Some students with LD—approximately 8% in this study—perform significantly less well in a read-aloud condition, for reasons which are as yet unclear. Though blanket assignment of students to an oral accommodation would likely improve group performance, this improvement would come at the expense of the small minority of students for whom a different form of construct-irrelevant variance is being introduced, resulting in scores that do not fully reflect their competence. Thus, the decision to assign students to an accommodated testing condition should only be made based on prior empirical evidence indicating the likelihood that the student's performance will benefit, or at least will not be impaired, in the accommodated condition.

A final consideration is that accommodations are not a remedy for low levels of skill on the construct that is being assessed. In the present study, students were not tested on grade-level mathematics items, but rather on items reflecting their current level of skill. This was necessary to detect the effect of the oral accommodation. However, if students in the present study had been tested at grade level, many may have performed poorly, even with an oral accommodation.

As additional studies clarify the aspects of test-taking situations that either improve or detract from student performance, a better understanding will also emerge of the skill sets—and accommodations—that students with and without disabilities may need to effectively use the assessed abilities in real-life contexts outside of school.

AUTHOR'S NOTE

The author gratefully acknowledges the collaboration of Dr. Maria Elena Arguelles in developing the mathematics tests used in this study.

REFERENCES

Calhoon, M. B., Fuchs, L. S., & Hamlett, C. L. (2000). Effects of computer-based test accommodations on mathematics performance assessments for secondary students with learning disabilities. *Learning Disability Quarterly, 23*(4), 271–282.

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

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Dale, E., & O'Rourke, J. (1979). *The living word vocabulary: The words we know.* Elgin, IL: Field Enterprises Educational Corp.

Edgemon, E., Jablonski, B., & Lloyd, J. (2006). Large-scale assessments: A teacher's guide to making decisions about accommodations. *Teaching Exceptional Children, 38*(3), 6–11.

Elbaum, B., Arguelles, M. E., Campbell, Y., & Saleh, M. B. (2004). Effects of a student-reads-aloud accommodation on the performance of students with and without learning disabilities on a test of reading comprehension. *Exceptionality, 12*(2), 71–87.

Elliott, S. N., Kratochwill, T. R., & McKevitt, B. C. (2001). Experimental analysis of the effects of testing accommodations on the scores of students with and without disabilities. *Journal of School Psychology, 39,* 3–24.

Harcourt Educational Measurement. (1996). *Stanford achievement test* (9th ed.). San Antonio, TX: Author.

Fuchs, L. (2000). *The validity of test accommodations for students with learning disabilities: Differential item performance on reading and mathematics tests as a function of test accommodations and disability status.* Wilmington: Delaware Department of Education.

Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., & Karns, K. M. (2000). Supplementing teacher judgments of mathematics test accommodations with objective data. *School Psychology Review, 29,* 65–85.

Helwig, R., Rozek-Tedesco, M., & Tindal, G. (2002). An oral versus a standard administration of a large-scale mathematics test. *The Journal of Special Education, 36*(1), 39–47.

Helwig, R., Rozek-Tedesco, M., Tindal, G., Heath, B., & Almond, P. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *The Journal of Educational Research, 93*(2), 113–125.

Helwig, R., & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children, 69,* 211–225.

Huynh, H., Meyer, P., & Gallant, D. J. (2004). Comparability of student performance between regular and oral administrations for a high-stakes mathematics test. *Applied Measurement in Education, 17,* 39–57.

Individuals with Disabilities Education Act Amendments of 1997, 20 U.S.C. 1400 *et seq.* (1997)

Individuals with Disabilities Education Improvement Act of 2004. P.L. 108–446. Retrieved January 16, 2006, from: http://frwebgate.access.gpo.gov/cgibin/getdoc.cgi?dbname=108_cong_public_laws&docid=f:publ446.108

Johnson, E. (2000). The effects of accommodations on performance assessments. *Remedial and Special Education, 21,* 261–267.

Johnson, E., Kimball, K., Brown, S. O., & Anderson, D. (2001). A statewide review of the use of accommodations in large-scale, high stakes assessments. *Exceptional Children, 67,* 251–264.

Meloy, L., Deville, C., & Frisbie, D. (2000). The effect of a read aloud accommodation on test scores of students with and without a learning disability in reading. *Remedial and Special Education, 23,* 248–255.

No Child Left Behind Act of 2001, 20 U.S.C. 70 § 6301 *et seq.* (2002)(Reauthorization of the Elementary and Secondary Education Act of 1965)

Phillips, S. E. (1994). High stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education, 7*(2), 93–120.

Pomplun, M., & Omar, M. H. (2000). Score comparability of a state mathematics assessment across students with and without reading accommodations. *The Journal of Applied Psychology, 85*(1), 21–29.

Schulte, A. A., Elliott, S. N., & Kratochwill, T. R. (2001). Effects of testing accommodations on students' standardized mathematics test scores: An experimental analysis. *School Psychology Review, 30,* 527–547.

Sireci, S., Scarpati, S., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research, 75*(4), 457–490.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved January 16, 2006, from: http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html

Thompson, S. J., Johnstone, C. J., Thurlow, M. L., & Altman, J. R. (2005). *2005 state special education outcomes: Steps forward in a decade of change.* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved January 11, 2006, from: http://education.umn.edu/NCEO/OnlinePubs/2005StateReport.htm/

Thurlow, M., & Bolt, S. (2001). *Empirical support for accommodations most often allowed in state policy* (Synthesis Report 41). Minneapolis: University of Minnesota National Center on Educational Outcomes.

Thurlow, M., House, A., Scott, D., & Ysseldyke, J. (2000). Students with disabilities in large-scale assessments: State participation and accommodation policies. *The Journal of Special Education, 34*(3), 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* (Technical Report 26). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Tindal, G. (2002). *Accommodating mathematics testing using a videotaped, read-aloud administration.* Washington, DC: Council of Chief State School Officers.

Tindal, G., & Fuchs, L. (1999). *A summary of research on test changes: An empirical basis for defining accommodations.* Unpublished manuscript, 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*(4), 439–450.

Weston, T. J. (2002). *The validity of oral accommodation in testing* (NCES 200306). Washington, DC: National Center for Education Statistics.