Exceptional Children

Vol. 69, No. 2, pp. 211-225.

An Experimental Analysis of Accommodation Decisions on Large-Scale Mathematics Tests

ROBERT HELWIG
GERALD TINDAL
University of Oregon

ABSTRACT: *This article reports on an investigation of teachers and students within special education to determine the accuracy with which teachers recommend read-aloud accommodations for mathematics tests, and develop a profile of students who benefit from this type of accommodation. Students in both general (*n = *973) and special education (*n = *245) in elementary and middle schools from eight states were administered an accommodated and standard mathematics achievement test. Teachers were no more successful than chance at predicting which students would benefit from the accommodation. Supplementary analyses used pretest reading and mathematics achievement scores in an attempt to develop a profile of students who favored one or the other formats. The outcomes from accommodations did not necessarily match student profiles.*

TEACHERS' ACCOMMODATION ASSIGNMENT

Due to recent mandates to include all students in large-scale testing (e.g., No Child Left Behind Act of 2002, IDEA 1997, and Improving America's Schools Act of 1994), increased attention is being paid to testing accommodations research for students with disabilities. In a comprehensive summary of this topic, Tindal and Fuchs (1999) cite 115 references from research on test accommodations. Over half of these were published in the 1990s. Thurlow, House, Scott, and Ysseldyke (2000) report that the number of states with active policies on accommodations increased from 21 in 1993 to 39 in 1997. Notwithstanding this interest, a lack of consensus exists over when accommodations are appropriate (Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000). One commonly used accommodation for students with low reading skill is the verbatim oral presentation of mathematics items. Limited research suggests that teachers are not effective at assigning this accommodation. The present article attempts to add empirical evidence to the field by examining the accuracy of teachers' perceptions of the effectiveness of read-aloud accommodations for students with disabilities. We also attempt to develop a profile of students who might benefit from this type of accommodation.

Importance of Accommodation Decisions

A testing accommodation is defined as a change in test presentation or response format that does not alter the construct under consideration (Tindal & Fuchs, 1999). The significance of accom-

modation decisions is evident in the consequences of poor choices. For example, certain types of accommodations give specific students access to tests otherwise unavailable to them. "Disallowing valid accommodations prevents students with disabilities from demonstrating their competence" (Fuchs, Fuchs, Hamlett, et. al, 2000, p. 68). One illustration of this is the practice of reading math test items aloud to students with low reading ability. Several studies indicate that this procedure differentially aids students with reading difficulties as compared to more able readers (Fuchs, Fuchs, Eaton, et al., 2000; Helwig, Rozek-Tedesco, Tindal, Heath, & Almond, 1999; Johnson, 2000; Tindal, Heath, Hollenbeck, Almond, & Harniss, 1998; Weston, 1999). For this accommodation to be useful, however, special education teachers, individualized education program (IEP) team members, and others involved in assessment decisions, must be accurate in identifying students who would benefit. The research reported in this article attempts to quantify this efficiency.

Misapplying this accommodation may also be detrimental to some students. Significant numbers of teachers feel that reading test items aloud is frustrating and distracting for some students (Weston, 1999). Some empirical evidence supports this supposition. For example, Helwig, Rozek-Tedesco, and Tindal (2002) found that higher skilled readers performed better on a standard administration of a mathematics test, rather than when items were read aloud to them. Fuchs, Fuchs, Eaton, et al., (2000) found similar results with certain students on a read-aloud accommodation of extended text passages. The reasons for this are not clear, although students differ in their abilities to listen (Joshi, Williams, & Wood, 1998), remember (Swanson, Cochran, & Ewers, 1990), and construct auditory versus visual schema (Luger, Johnson, Stern, Newman, & Yeo, 1994; Oakhill & Yuill, 1996).

It is also possible that the background noise of oral presentations might be distracting for some students who choose to read items for themselves. Evidence suggests that the comprehension of skilled readers decreases when text is read aloud (Goldman, Hogaboam, Bell, & Perfetti, 1980). Accommodations tend to be overrecommended by teachers for students with disabilities (Fuchs, Fuchs, Eaton, et al., 2000; Fuchs, Fuchs, Hamlett, et al., 2000; Johnson, Kimball, Brown, & Anderson, 2001; Weston, 1999). This practice increases the likelihood that students will be adversely affected by an unneeded accommodation.

A further advantage of only administering accommodations to students who will benefit from them concerns the expenditure of valuable resources. Nonstandard test administrations are typically more costly in terms of time or money. Presenting test items orally requires districts to train staff in test administration procedures as well as schedule alternative testing sites and times for students requiring this type of accommodation. When states and schools are faced with limited resources, social consequences quickly become part of measurement validity.

Finally, accommodation decisions affect the validity of inferences made from test results. While policies vary from state to state, certain changes in test administration procedures may result in scores being disaggregated from scores derived under standard testing conditions. If the results of significant numbers of students are not included in data analyses because their testing conditions were unnecessarily altered, critical policy decisions may be based on incomplete information. Further, in these instances individual students and their teachers are denied valuable comparative achievement information because disaggregated scores may have a different meaning than scores obtained under standard conditions.

Teacher Effectiveness at Assigning Accommodations

Only two studies have examined the accuracy of teacher perception of the effectiveness of read-aloud accommodations for mathematics tests. These offer conflicting results. In an investigation of fourth-grade students and their teachers, Weston (1999) noted from interviews that "*All* teachers felt that reading content in test items

When states and schools are faced with limited resources, social consequences quickly become part of measurement validity.

poses a barrier for their students, and expected very poor readers to do much better on the accommodated test than on the non-accommodated test" (p. 126; emphasis added). He found, however, that teachers did no better than chance when predicting which students would benefit from the accommodation.

Fuchs, Fuchs, Eaton, et al. (2000) presented a copy of a standardized mathematics test to a group of general and special education teachers and asked them to identify which of their students would benefit by having the items read aloud. Students subsequently were administered this test twice, once in standard format, and once with items read aloud. Results showed that on concept/application problems with limited reading demands, students whose teachers had recommended an accommodation benefited more from the read-aloud format than students whose teachers did not recommend this accommodation. On problem-solving items with extended reading requirements, however, recommended and nonrecommended students benefited equally from the accommodation.

In this latter study, it was found that when information concerning the level of effectiveness of previous accommodations was available, the efficiency of sorting students could be increased for both concepts/applications as well as problem-solving tasks. (For more information on this diagnostic procedure called the Dynamic Assessment of Testing Accommodations [DATA] see Fuchs, Fuchs, Eaton, et al. [2000].) Another system of increasing teachers' efficiency in recommending accommodations is the Assessment Accommodations Checklist [AAC] Elliott, Kratochwill, & Schulte (1998) designed to "help organize, record, and encourage teachers to evaluate the helpfulness and fairness of assessment accommodations provided students" (p. 11). Both systems rely on increasing the amount of information available to decision makers.

Student Profiles

Only general information is known about which students benefit from read-aloud accommodations for mathematics tests. Specific conclusions are difficult because of the different comparison populations used in previous studies. These included students within special education, students with learning disabilities (LDs), students with IEPs in reading, low-skilled readers, or combinations of reading and math skill. Confounding these grouping strategies is the fact that many students within these target groups did not benefit while many students in control groups did.

A few studies give some clues as to which students might be candidates for oral reading accommodations. Helwig et al. (1999) found that sixth-grade students with low reading proficiency but average and above math skills benefited from a reading accommodation, whereas able readers as well as low readers with poor math skills did not. Fuchs, Fuchs, Eaton, et al., (2000) found that among fourth-grade students with LD, those more proficient in math skills benefited more from a read-aloud accommodation than their lower achieving counterparts. Weston (1999) found the correlations between reading achievement and the effects of a read-aloud accommodation for fourth-grade students were significant for students with disabilities but not for others. These findings suggest that there may be both a reading proficiency threshold above which accommodations are no longer effective, as well as a math threshold below which students do not have the skills necessary to take advantage of the accommodation.

In summary, testing accommodations are a valuable tool to ensure that test results and their interpretations are valid and fair. The inappropriate use or withholding of accommodations, however, can invalidate scores, rob students and educators of valuable information, and increase costs. It appears that teachers, in many cases, are not efficient in recommending which students will benefit from having math items read aloud. Organized systems such as DATA or AAC may increase this efficiency. In the absence of this type of data concerning previous accommodations, however, it appears that students' knowledge of math and reading proficiency levels might prove useful in assigning accommodations. The purpose

It appears that teachers, in many cases, are not efficient in recommending which students will benefit from having math items read aloud.

of the current research was twofold. First we examined teacher accuracy in predicting accommodation effectiveness over four grade levels with students with disabilities to see if previous research results were confirmed. Next we examined student pretest achievement scores in reading and math to determine whether patterns of proficiency in these basic skill areas could be used to develop a profile of students likely to benefit from read-aloud accommodations.

METHOD

As part of a larger study, state departments of education in eight states were requested to identify districts that would be interested in participating in an accommodations project. Subsequently, one middle school and one elementary school were identified within each state. Each school agreed to the participation of two general education classrooms (25-30 students) with a common grade level (fourth- or fifth-grade in elementary school or seventh- or eigth-grade in middle school). In addition, within each school, all students at the chosen grade level receiving services within special education, and for which large-scale mathematics testing was part of their academic program, were also included. For the purposes of the present research, except where noted, only these latter students were included in data analyses.

Subjects

Students. A total of 1,550 students took part in this project, of which 1,218 completed all measures. Primary data analyses were conducted for 245 students receiving special education services. Over 70% of these students were identified with learning disabilities. The next most common disabilities were speech or language impairments (8%), serious emotional disturbance (5%), and mental retardation (5%). Males outnumbered females 63% to 37%. The majority of students were White (59%) with Blacks comprising 28% of the population, and Hispanics 7%. Demographic information, broken down by grade level, is presented in Table 1.

Teachers. The teachers included in this study were those providing the mathematics instruction for the students described above. In some cases this was a special education teacher, while in others, it was a general education teacher. In those instances where students received instruction in multiple locations, survey information (described below) was completed by the person(s) with the greatest knowledge of the student's mathematical achievement and ability levels. It was acceptable for special and general education teachers to collaborate on an individual student's survey questionnaire if they possessed distinct knowledge of the students' abilities. Due to this data collection method, it was impossible to determine the exact identity of the teacher who completed any specific survey item. It was known, however, that the 245 students under investigation came from 61 different classrooms.

Measure

Math Achievement. Two 30-item, multiple-choice, mathematics achievement tests (Form A and Form B) were created from a secured pool of approximately 100 items from a participating state's seventh-grade item bank. All items were reviewed by department of education officials within the remaining seven states for congruence with their state's curriculum. Items utilizing subject matter not covered by any state's seventh-grade curriculum were discarded from the pool. The 30-item tests were matched for content and difficulty based on field-testing data from the home state. Each test covered a variety of domains including algebraic relationships, measurement, statistics and probability, and geometry. Items consisted of word problems of 7 to 45 words including the four answer choices. No straight calculation problems were included. On each version, a student's score was the total number of items answered correctly. These tests were used with all seventh- and eigth-grade participants.

Each form (A and B) was created in two formats. The first was standard format with items presented in written form in a test booklet with several items per page. In addition, a video was created for each test form showing the face of an actor reading each item on one portion of the screen while the text of the problem was shown on the remaining portion. A test booklet with one item per two facing pages was included with this format. In both formats, students bubbled their answers on a separate answer sheet.

TABLE 1
Demographic Information

Group	Grade Level			
	4	5	7	8
Total	506	277	450	317
Completed all measures	413	224	316	265
Special education	63	41	62	79
Gender				
Male	37	26	38	52
Female	25	15	24	27
Missing	1	0	0	0
Ethnicity				
White	37	23	35	50
African American	12	16	14	26
Hispanic	12	0	5	1
Asian/Pacific Islander	0	0	0	1
American Indian	1	0	4	1
Multiracial	1	1	4	0
Missing	0	1	0	0
Disability				
Learning disability	52	29	41	54
Speech/ language impairment	11	4	4	1
Emotional disturbance	2	2	3	7
Mental retardation	1	0	6	6
Orthopedic impairment	0	2	0	1
Hearing impairment	1	0	1	1
Visual impairment	2	0	0	0
Other	0	3	3	6
Missing	6	2	11	10

Note: Disability may sum to greater than 100% due to multiple disabilities.

An identical procedure was followed with a starting pool of approximately 100 items taken from the same state's fourth-grade item bank. The two resulting tests were used with all elementary students. The domains covered were identical to those described above. The number of words per item, including answer choices, for these tests ranged from 7 to 69.

Teacher Ratings. For each student, the appropriate teacher (as described earlier) completed a survey that rated the student's skill level in both reading and mathematics on a 5-point Likert scale ranging from *very low proficiency* to *very high proficiency*. The survey also posed the following question, "How important in generating successful performance is the accommodation (video version) for this student?" Respondents were given five choices ranging from *very low importance* to *very high importance*.

Basic Math Skills Test. A math skills test was given to all students. In the middle school, this test consisted of 21 items, 16 of which were basic skills computations involving addition, subtraction, multiplication, and division of whole numbers, fractions, and decimals. Three items involved transforming decimals, fractions, and percents into a different form. The final two items involved one-step word problems utilizing 20 or fewer words. All items were production response. A previous field test involving 240 middle school students resulted in a correlation of .75 between the total number correct on this test and a 60-item multiple-choice problem-solving test used as part of a southwestern state's statewide math assessment program.

The math skills test for fourth- and fifth-grade students consisted of 17 addition, subtraction, multiplication, and division items involving whole numbers and fractions. Two additional problems utilizing less than seven words required students to convert or add units of measure. In a field test, the correlation between this task and the 60-item elementary version of the standardized math achievement test used in the present

study was .60. The use of calculators was not permitted for either the elementary or middle school versions of this test.

Reading Maze. Two maze tasks, one for elementary students and one for middle school students, were used as measures of reading proficiency. These consisted of passages of approximately 200 words with 25 of these replaced with blanks. For each blank, students were required to select, from five choices, the word that had been deleted. Unlike typical maze tasks in which every nth word is omitted, words judged important to the story were chosen. In all cases these were nouns, verbs, adjectives, or adverbs. Previous field tests revealed correlations of .68 (elementary) and .76 (middle school) between these mazes and total scores from 60-item reading comprehension tests used as part of a statewide assessment program. In the current study, correlation coefficients between the maze and the math tasks ranged from .55 to .60 in the elementary school and .41 to .46 in the middle school.

Procedures

At both elementary and middle school grade levels, classrooms of students were randomly assigned to two groups. Both groups took each form (A and B) of the test with 1- to 4-day intervals between administrations. Group 1 took Form A in standard format and Form B in video format. Group 2 took Form A in video format and Form B in standard format. Within each group, the order of administration (Forms A and B) was counterbalanced. All tests were group-administered.

When completing the standard version of the tests, students were given a test booklet and instructed to read each item, choose the best answer from the answer choices, and complete the test at their own pace. No time limit was set. For the video presentation, students were given a choice of watching and listening to the items read and displayed on the video monitor or following along with the reader as they silently read from their booklet. Test administrators controlled the video display. At the completion of the reading of each item, the video was paused for a predetermined length of time depending on problem type. During this period, students worked on the solution to each problem and marked their choices. After the designated pause, students were instructed to turn the page, at which time another problem was displayed. Test booklets were provided so that students could reread an item or scan it for critical information. Students were instructed to not turn the pages of their booklets until the next problem was presented on the monitor. The video tests were group-administered in classrooms with seating arranged so that all students had a clear view of the monitor.

Based on field testing, suggested pause times between problems were set from 15 to 60 seconds and were considered liberal. As an added protection, test administrators were instructed to use their discretion if they felt some students could not complete a problem within the recommended timeframe. In these cases, additional time was permitted. This precaution was designed to alleviate time pressures on students to complete problems and helped to ensure that our administrator-paced format would more closely mirror the student-paced format that is typical of most testing accommodations of this type. In no instances were pause times allowed to be shortened, even if it appeared that all students had completed an item.

DATA ANALYSES

For all Group 1 elementary students (general and special education), we calculated standard (z) scores for both the 30-item paper-and-pencil and 30-item oral presentation tests. We subtracted the paper-and-pencil z-score from the accommodated z-score, giving us a measure of the magnitude of the difference (Diff) in relative performance between the two formats in standard deviation units. A positive number indicated that a student's strongest relative performance had occurred with the oral presentation. We noted each student with a difference greater than or equal to the absolute value of .5. Thus, we had a list of all Group 1 students who experienced a shift, in either direction, of at least one-half of a standard deviation between the standard and accommodated formats of the tests. We deemed this a noticeable shift with practical implications. These students were labeled either "Favor Standard" or "Favor Oral." We repeated this exact procedure for middle school students, and then again for Group 2

Implementing an accommodation within the classroom prior to testing is likely to benefit students.

members. The majority of the data analysis discussed in this article includes only those students within special education who experienced this .5 magnitude shift.

To investigate the accuracy with which teachers assigned read-aloud accommodations, we separated students into four groups based on (a) their strongest test performance (standard or video), and (b) their teachers' belief in the need for a test accommodation. We based this latter criterion on teachers' responses to our survey. We considered a student recommended for an accommodation if the teacher had marked the accommodation as *high importance* or *very high importance*. We considered a student not recommended for an accommodation if the teacher had marked the accommodation as *low importance* or *very low importance*. Two of the groups represented students whose teachers' predictions had been accurate (recommended students who performed better in the video format and not recommended students who performed better in the standard format). The remaining two groups represented inaccurate predictions (recommended students who performed better on the standard format and not recommended students who performed better on the video version). The percentage of correct predictions was used as an indicator of teacher accuracy.

To develop a profile of students who benefited from read-aloud accommodations, we used *t*-tests to compare the basic math skill and maze scores of the Favor Standard and Favor Oral groups. Further, to investigate whether math and reading skills worked in conjunction we divided students into four groups, first by reading level (using a Maze *z*-score of -0.5 as the demarcation between poor and satisfactory reading) and then by basic math skill (using a math skill *z*-score of -0.5 as the dividing line). Thus, we identified poor readers with both adequate and inadequate math skills as well as competent readers with both adequate and inadequate math skills. A relatively low standard was set to differentiate between high and low achievers for two reasons. First, we felt that students with math and reading achievement scores near the mean (*z*-score = 0) would not normally receive testing accommodations. Thus, our terms *high* and *low* are not used in a normative sense, but rather, are used in relation to our specific purpose. Second, the criteria we chose enabled us to include a sufficient number of students in each category to perform meaningful statistical analyses. Following the grouping of students, we used analysis of variance (ANOVA) to compare the Diff scores of each group.

RESULTS

Complete data sets (survey information, maze, basic math skills test, and both formats of the mathematics achievement test) were collected for 1,218 of the original 1,550 students. Means and standard deviations for all achievement variables, split by student classification and grade level, are presented in Table 2. Unpaired *t*-tests showed a significant difference ($p < .01$) between general and special education students on all measures within all grade levels. Table 3 shows the distribution of teacher ratings of students' needs for accommodations. The importance of an accommodation was rated as high or very high for approximately 56% of students within special education. In contrast, only 9% of these students received a low or very low importance rating.

Almost 600 (46%) of the students experienced a shift of a least one-half of a standard deviation between the paper-and-pencil and video presentations of the mathematics achievement test. These were approximately evenly split between the two testing formats. Among students in special education experiencing a shift of this magnitude ($n = 122$), 62 favored the standard version and 60 favored the video version.

Teacher Accuracy

Table 4 shows teacher ratings of accommodation importance for students in special education rated as either needing (rating 4 or 5) or not needing (rating 1 or 2) an accommodation.

These are broken down by student actual performance preference. To increase cell size for data analysis, fourth- and fifth-grades were combined as were seventh- and eighth-grades, to form elementary and middle school categories. Of the 122 students who favored one or the other testing formats, 80 were rated at the extremes (ratings of 3 were eliminated) with a strong majority (81%) of these students rated as needing an accommodation.

In elementary school, 52% of students whose teachers felt it was important for them to receive a test accommodation performed better under standard conditions. These students represented incorrect recommendations. In middle school, 50% of students represented incorrect recommendations. For elementary students whose teachers felt an accommodation was of low importance, 14% performed better on the video test, whereas 88% of middle school students whose teachers felt an accommodation was not important performed better on the video test. These students also represent incorrect recommendations. Overall, 45% of elementary students and 57% of middle school students were incorrectly recommended by their teachers.

Profile

Table 5 compares the mean maze and basic math skill scores for the Favor Standard and Favor Oral groups. There were three significant differences in the mean scores of these variables when comparing students who did better on the standard test versus those who did better on the accommodated version. In the fourth grade, students who did better in the accommodated format had significantly higher pretest scores on both the maze and math skills test. In the fifth grade, however, we found the opposite result. Students who did better on the standard version had significantly higher math skill scores, and their higher maze scores approached significance (p = .09). There were no significant differences in middle school.

Eliminating a novelty effect by putting accommodations into daily classroom practice might increase the accuracy of teacher judgments.

Our final analysis (Table 6) compared the Diff scores (z-score differences between standard and accommodated formats) of the four achievement groups (LowRead/LowMath, HiRead/LowMath, LowRead/HiMath, and HiRead/HiMath). At both levels, the largest mean (in absolute value) occurred within the LowRead/HiMath group. In each case the average performance was approximately one-half of a standard deviation in favor of the standard test administration. ANOVA revealed a main effect for group in both elementary (F [3,100] = 4.36) and middle school (F [3,137] = 6.08). For elementary students, follow-up contrasts (Scheffe's F) revealed a significant difference between LowRead/HiMath versus HiRead/LowMath. In middle school, two of the comparisons were significant: LowRead/HiMath versus HiRead/LowMath and LowRead/HiMath versus LowRead/LowMath.

DISCUSSION

This research posed two questions. How accurate are teachers' recommendations of read-aloud accommodation for math tests? And what is the profile of a student who benefits from this type of accommodation? Unfortunately, we were only able to formulate a tentative answer to one of these inquiries.

Teacher Prediction Accuracy

The teachers in our study did not appear to be effective in their recommendations of which students would, and would not, benefit from having math test items read aloud. Teachers' ratings of their students' needs for testing accommodations coincided with actual student performance only approximately half of the time. This result is no better than what would be expected by simple random guessing.

Consistent with the work of Fuchs and colleagues (Fuchs, Fuchs, Eaton, et al., 2000; Fuchs, Fuchs, Hamlett, et al., 2000), we found that teachers tend to overrecommend accommoda-

TABLE 2
Means and Standard Deviations for Achievement Variables

	Form A						*Form B*									
	Standard			*Accommodated*			*Standard*			*Accommodated*			*Maze*		*Math Skill*	
Group	n	M	SD	n	M	SD	n	M	SD	n	M	SD	M	SD	M	SD
Grade 4																
Total	163	16.64	6.67	250	17.55	6.90	250	16.96	6.30	163	16.88	5.82	20.56	5.46	12.95	2.87
General Ed.	141	17.62	6.25	209	18.78	6.51	209	18.00	6.01	141	17.44	5.72	21.80	4.07	13.49	2.43
Special Ed.	22	10.41	5.96	41	11.29	5.25	41	11.71	5.04	22	13.27	5.23	13.65	6.87	9.95	3.28
Grade 5																
Total	181	20.31	6.52	43	25.23	4.00	43	23.42	5.01	181	20.39	5.50	22.09	3.85	14.82	2.54
General Ed.	146	21.89	5.42	37	26.00	3.29	37	24.16	4.29	146	21.77	4.84	22.84	2.84	15.32	2.22
Special Ed.	35	13.74	6.65	6	20.50	5.01	6	18.83	6.97	35	14.66	4.32	18.73	5.65	12.01	2.70
Grade 7																
Total	171	12.78	5.58	145	12.26	5.34	145	12.56	4.72	171	12.66	5.46	20.03	5.13	13.12	4.38
General Ed.	139	13.80	5.55	115	12.97	5.57	115	13.19	4.85	139	13.93	5.11	20.91	4.57	14.04	3.99
Special Ed.	32	8.34	2.92	30	9.53	3.16	30	10.13	3.22	32	7.16	2.95	16.40	5.70	9.35	3.91
Grade 8																
Total	123	15.48	5.85	142	13.31	4.96	142	13.56	5.25	123	15.70	5.30	21.05	4.33	13.38	4.42
General Ed.	100	16.71	5.53	86	14.92	4.68	86	15.63	5.07	100	16.92	4.87	22.49	2.31	15.16	3.27
Special Ed.	23	10.13	3.86	56	10.84	4.34	56	10.39	3.76	23	10.39	3.60	17.63	5.84	9.20	3.95

TABLE 3
Frequency Distribution of Teacher Ratings of Student Needs for Accommodations

	Very Low		*Low*		*Fair*		*High*		*Very High*	
Group	n	%	n	%	n	%	n	%	n	%
All Grades										
Total	164	13	262	22	451	37	266	22	75	6
General Ed.	161	17	243	25	365	38	178	18	26	3
Special Ed.	3	1	19	8	86	35	88	36	49	20
Grade 4										
Total	34	8	100	24	160	39	97	23	22	5
General Ed.	34	10	92	26	149	43	67	19	8	2
Special Ed.	0	0	8	13	11	17	30	48	14	22
Grade 5										
Total	51	23	39	17	73	33	46	21	15	7
General Ed.	50	27	37	20	58	32	32	17	6	3
Special Ed.	1	2	2	5	15	37	14	34	9	22
Grade 7										
Total	54	17	68	22	118	37	54	17	22	7
General Ed.	53	21	62	24	89	35	39	15	11	4
Special Ed.	1	2	6	10	29	47	15	24	11	18
Grade 8										
Total	25	9	55	21	100	38	69	26	16	6
General Ed.	24	13	52	28	69	37	40	22	1	1
Special Ed.	1	1	3	4	31	39	29	37	15	19

TABLE 4
Summary of Teacher Accommodation Recommendation by Actual Student Performance

	Teacher Accommodation Recommendation	
Best Performance	*No Need*	*Need*
Elementary School		
Standard	6	16
Accommodated	1	15
Middle School		
Standard	1	17
Accommodated	7	17

tions. This is not surprising. It makes intuitive sense that students who have difficulty reading text for themselves would benefit by having the text read aloud for them. The teachers in Weston's (1999) study were unanimous in this perception. In fact, however, only approximately half of the students in our study improved their relative performance when tests items were read aloud, and only about half of these students increased their performance by at least one-half of a standard deviation. This was the case with students in both general and special education. Overall, it did not make any difference in which direction teachers felt strongly about students' accommodations. Teachers' perceptions of the importance of accommodations for individual students were unrelated to the actual benefit students received.

One limitation of our study is that we purposefully excluded from our analyses those students who did not experience a change in performance of at least one-half of a standard deviation between the standard and accommodated formats. This represents approximately half of the students tested. Thus, we may have ignored many correct teacher recommendations, resulting in a more pessimistic view of teacher accuracy than we would have otherwise obtained. While this may be the case, we defend our analyses methods based on two factors.

TABLE 5
Comparison of Maze and Math Basic Skills Scores by Favored Format

	Favor Standard					*Favor Oral*							
		Maze		*Math Skill*			*Maze*		*Math Skill*		*Maze*		*Math Skill*	
Group	n	M	SD	M	SD	n	M	SD	M	SD	*t-value*	*p-value*	*t-value*	*p-value*
All	62	15.44	7.04	9.48	4.40	60	16.12	6.22	10.17	3.69	-.57	.57	-.93	.36
Elementary	26	13.69	7.29	10.38	3.71	27	15.85	6.22	11.30	2.45	-1.16	.25	-1.06	.29
Grade 4	18	10.50	6.06	8.61	2.79	14	15.43	6.09	11.50	3.06	-2.28	.03	-2.79	.01
Grade 5	8	20.88	3.94	14.38	2.00	13	16.31	6.58	11.08	1.66	1.77	.09	4.10	.001
Middle School	36	16.69	6.67	8.83	4.78	33	16.33	6.31	9.24	4.27	.23	.82	-.37	.71
Grade 7	19	17.05	6.43	9.74	4.78	11	15.18	4.85	8.18	3.63	.83	.41	.93	.36
Grade 8	17	16.29	7.10	7.82	4.72	22	16.91	6.96	9.77	4.55	-.27	.79	-1.31	.20

TABLE 6
Means and Standard Deviations for Difference by Group Membership

	Elementary School			*Middle School*		
Group	n	M	SD	n	M	SD
LowRead/LowMath	47	.07	.62	55	.10	.59
HiRead/LowMath	23	.33	.56	33	.14	.60
LowRead/HiMath	10	-.45	.88	16	-.51	.63
HiRead/HiMath	24	-.12	.51	37	-.26	.71

Note: Position means indicate performance favored the accommodated version.

First, because there is relatively little work in this area, it is reasonable that we initially focused on those students who stood to gain the most. Acquiring information on how accurately teachers perceive the effectiveness of reading test items aloud for students significantly helped or hindered by this format is clearly of practical use. There is limited practical value in examining teacher accuracy as it pertains to students who make relatively small changes in relative performance. Second, our method gave teachers an advantage in predicting performance. If teachers have insight into how students will perform with test accommodations, surely this insight would be more pronounced for individuals at the extremes. It seems clear that it would be easier to guess which format a student would prefer if the pool only contained subjects with clear preferences. The fact that teachers were not effective in their predictions, even for students at the extremes, makes a strong statement about teacher perceptions.

Student Profile

We are careful, however, not to criticize teachers' apparent lack of insight into their students' needs for read-aloud accommodations. Given the benefit of post facto data and statistical analysis techniques, we were unable to do much better. We attempted to develop a profile of those students who benefit from read-aloud accommodations by contrasting their achievement levels in reading and basic math skills with students who performed better on the standard format of the test. Surprisingly, we found little difference between the two groups. Only 3 of 14 comparisons we made were significant, and these offered conflicting information. The fourth-grade students who favored the accommodated format scored higher on both math and reading measures. Similar high achievers in the fifth grade, however, favored the standard format. In middle school, we found no relationship between students' math or reading performance and their favoring of a particular testing format.

Teachers' ratings of their students' needs for testing accommodations coincided with actual student performance only approximately half of the time.

Reasoning that math and reading skills might work in conjunction to influence accommodation effectiveness, we placed individuals in four groups. Though we found some significant differences, they were not in the anticipated direction. We had expected that students with low reading skills but adequate math skills would be the most likely to take advantage of a read-aloud accommodation. These appeared to be the students who would have difficulty deriving meaning from text, but at the same time their math achievements levels indicated that once they comprehended the problem statement, they would possess the mathematical tools necessary to solve many of the problems.

Contrary to our hypothesis, in both elementary and middle school, the low readers with adequate math skills performed better on the standard version by an average of one-half of a standard deviation. This is surprising, not only because it is counterintuitive, but because Helwig et al. (1999) found that these students were the only group to favor a read-aloud accommodation. The fact that these researchers limited their examination to six test items meeting complex linguistic criteria may explain the effect they found. The items used in this previous study were generally of a higher read-

Contrary to our hypothesis, in both elementary and middle school, the low readers with adequate math skills performed better on the standard version by an average of one-half of a standard deviation.

ability level than the ones used in the current study. It is not surprising that students receive more benefit from read-aloud accommodations when the readability level of items is high than when items contain relatively simple language. It may be the case that significant numbers of students were able to gain sufficient solution strategy clues from key words or by recognizing a problem type without having to read an item in its entirety. This does not explain, however, why we found that students performed better when they had to read items themselves. If it is true that some readers were distracted by our video presentation, it seems most reasonable that it would be skilled readers as reported by Goldman et al. (1980).

Another counterintuitive finding was the fact that in both elementary and middle school the performance difference of the LowRead/HiMath group (Favor Standard) differed significantly from the HiRead/LowMath group (Favor Oral). We did not anticipate that our accommodation would offer anything to this latter group of able readers with low math skills. This is precisely the group least likely to benefit from having items read aloud (high reading skill) and least likely to apply the information (low math skill). Replication studies are needed to determine if these patterns hold for other populations.

IMPLICATIONS FOR PRACTICE

Our research confirms the findings of others that teachers are not accurate in their assignment of accommodations. Nevertheless, the importance of accommodation decisions for students in special education remains, considering the need for validity in testing and federal mandates to include all students in large-scale testing programs. Because teachers are the individuals working closest with students, and possessing the greatest knowledge of each student's capabilities, it is the job of researchers to develop methods to increase teacher efficiency rather than to bypass them in the decision-making chain.

Systems such as DATA (Fuchs, Fuchs, Eaton, et al., 2000; Fuchs, Fuchs, Hamlett, et al., 2000), which target students for accommodations based on results from previous accommodations, hold promise and warrant our tentative recommendation. The vast majority of teachers, however, do not have access to these types of systems and must rely on other methods. Previous research suggested that the use of math and reading screeners might be useful for identifying individuals as possible candidates for read-aloud accommodations. Unfortunately, the present study did not confirm this. We do not completely abandon this method, however. Used as the primary means of identification, this system seems to have limitations. As an alternative, it may be the case that teachers or testing coordinators could use their intimate knowledge of students (unavailable to us) in conjunction with math and reading profiles to enhance decision-making efficiency. For example, we defined math and reading skill levels based on achievement test results. Reading disabilities, however, come in a variety of forms. While we might be tempted to avoid recommending a read-aloud accommodation to an individual scoring high on a reading maze screener, students exist who possess high reading comprehension (one of the traits measured by maze tests) as well as low fluency. If firsthand teacher knowledge of a student in this category revealed that he or she became easily frustrated when reading difficult text, it might be advisable to implement a read aloud accommodation even though the reading screener suggested that this was not necessary.

Because teachers are the individuals working closest with students, and possessing the greatest knowledge of each student's capabilities, it is the job of researchers to develop methods to increase teacher efficiency rather than to bypass them in the decision-making chain.

Implementing an accommodation within the classroom prior to testing is also likely to benefit students. While this is a prerequisite in some states, many states do not require an assessment/instruction link. Math and reading screening information, as well as firsthand teacher knowledge, may indicate that a particular student would benefit from having test items read aloud. However, the unfamiliarity of this testing format may prove sufficiently distracting or confusing to negate any positive effects that might be expected. In an effort to eliminate one irrelevant testing factor (reading ability), implementing an accommodation students have never seen before may introduce another. This is one possible explanation for the ineffectiveness of the two methods we evaluated. Eliminating a novelty effect by putting accommodations into daily classroom practice might increase the accuracy of teacher judgments.

Our inability to validate an effective identification system should not discourage teachers from assigning this type of accommodation with the proper precautions. There is only limited data to suggest that over-assigning read-aloud accommodations is harmful to students. Moreover, the potential harm is likely greatest for students not typically targeted for accommodations (for example, high-ability readers). Until empirical research validates a practical system for assigning read-aloud accommodations, teachers and IEP team members should rely on their knowledge of students' reading and mathematics achievement, learning styles, classroom experiences, and testing behaviors. With current and relevant information in all of these areas, teachers should be able to more accurately target students for accommodations.

REFERENCES

Elliott, S. N., Kratochwill, T. R., & Schulte, A. G. (1998). The assessment accommodation checklist: Who, what, where, when, why, and how. *TEACHING Exceptional Children, 31*(2), 10-14.

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 sources. *School Psychology Review, 29*, 65-85.

Fuchs, L. S., Fuchs, D., Hamlett, C., Eaton, S. B., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments 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*, 647-655.

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

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

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

Johnson, E. S., 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.

Joshi, R. M., Williams, K. A., & Wood, J. R. (1998). Predicting reading comprehension from listening comprehension: Is this the answer to the IQ debate? In C. Hulme & R. M. Joshi (Eds.), *Reading and spelling: Development and disorders* (pp. 319-327). Mahwah, NJ: Lawrence Erlbaum Associates.

Luger, G. P., Johnson, P., Stern, C., Newman, J. E., & Yeo, R. (1994). *Cognitive science: The science of intelligent systems*. San Diego, CA: Academic Press.

Oakhill, J., & Yuill, N. (1996). Higher order factors in comprehension disability: Processes and remediation. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and interventions* (pp. 69-92). Mahwah, NJ: Lawrence Erlbaum Associates.

Swanson, H. L., Cochran, K. F., & Ewers, C. A. (1990). Can learning disabilities be determined from working memory performance? *Journal of Learning Disabilities, 23*, 59-67.

Thurlow, M., House, A. L., Scott, D. L., & Ysseldyke, J. (2000). Students with disabilities in large-scale assessments: State participation and accommodation policies. *Journal of Special Education, 34*, 154-163.

Tindal, G., & Fuchs, L. S. (1999). *A summary of research on test change: An empirical basis for defining accommodation*. Lexington: University of Kentucky, Mid-South Regional Resource Center.

Tindal, G., Heath, B., Hollenbeck, K., Almond, P. J., & Harniss, M. (1998). Accommodating students with disabilities on large-scale tests: An experimental study. *Exceptional Children, 64*, 439-450.

Weston, T. J. (1999). Investigating the validity of the accommodation of oral presentation in testing (Doctoral dissertation, University of Colorado, 1999). *Dissertation Abstracts International, 60*, A1083.

ABOUT THE AUTHORS

ROBERT HELWIG, Research Associate; and **GERALD TINDAL**, Director, Behavioral Research and Teaching, University of Oregon, Eugene.

IN MEMORY

The research described in this article was an important part of Dr. Robert Helwig's life. On September 2, 2002, he died of a sudden heart attack. He will be remembered for his careful work and dedication to getting things right.

Correspondence concerning this article may be addressed to Gerald Tindal, Behavioral Research and Teaching, 5262 University of Oregon, Eugene, OR 97403-5262.

Manuscript received January 2002; accepted July 2002.

Provide your Students with the Necessary Skills Needed to Make a Successful Transition from School to Adult Life

A functional curriculum that provides community-based educational experiences is key!!!

Look to

Life Centered Career Education

Life Centered Career Education—LCCE—will help special education teachers meet the transition and testing requirements of IDEA 1997. It includes IEP goals and objectives and three types of assessment devices: rating scales, standardized knowledge batteries, and performance batteries. A highly engaging, motivating, and effective classroom, home and community-based curriculum, LCCE employs sound teaching practices to prepare students to function independently and productively as family members, citizens and workers, and to enjoy fulfilling personal lives.

LCCE is ideal for the late elementary grades through adult education. First developed for students with learning disabilities or mild developmental disabilities, it is also appropriate for students at risk, students for whom English is a second language, and students in general education. A modified guide to using the curriculum for students with moderate disabilities is available.

The Major Components

<u>Daily Living Skills</u>
Managing Personal Finances
Managing a Household
Exhibiting Responsible Citizen-

<u>Personal Social Skills</u>
Achieving Socially Responsible Behavior
Maintaining Good Interpersonal Skills
Acquiring Problem Solving Skills
And More....

<u>Occupational Skills</u>
Exploring Occupational Options
Planning Occupational Choices
Exhibiting Appropriate Work Habits
Seeking and Maintaining Employment and more

LCCE also includes IEP goals and objectives and three types of assessment devices: rating scales, standardized knowledge batteries, and performance batteries.

For more information on Life Centered Career Education
call (703) 264-9451 or e-mail **geneen@cec.sped.org**