**WCER Working Paper No. 2003-7**
**June 2003**

# Effects of Testing Accommodations on Math and Reading Scores: An Experimental Analysis of the Performance of Fourth- and Eighth-Grade Students With and Without Disabilities

**Ryan J. Kettler**
Department of Educational Psychology/
Wisconsin Center for Education Research
University of Wisconsin–Madison
rjkettler@students.wisc.edu

**Bradley C. Niebling**
Department of Educational Psychology/
Wisconsin Center for Education Research
University of Wisconsin–Madison
bcniebling@wisc.edu

**Andrew A. Mroch**
Department of Educational Psychology/
Wisconsin Center for Education Research
University of Wisconsin–Madison
aamroch@wisc.edu

**Elizabeth S. Feldman**
Department of Educational Psychology/
Wisconsin Center for Education Research
University of Wisconsin–Madison
esfeldman@wisc.edu

**Markeda L. Newell**
Department of Educational Psychology/
Wisconsin Center for Education Research
University of Wisconsin–Madison
mlnewell@wisc.edu

Wisconsin Center for Education Research
School of Education • University of Wisconsin–Madison
http://www.wcer.wisc.edu/

The research reported in this paper was supported by the U.S. Department of Education, Office of Special Education and Rehabilitative Services, in grants awarded to Stephen N. Elliott and Thomas R. Kratochwill on *Analysis of the Effects and Consequences of Testing Accommodations on Students' Achievement Test Scores* (#H324D990027) and to Stephen N. Elliott, Thomas R. Kratochwill, and Daniel Bolt on *Using DIF Analyses to Examine the Effects of Testing Accommodations on Students' Responses to Test Items* (#H324C000033), and by the Wisconsin Center for Education Research, School of Education, University of Wisconsin–Madison. Any opinions, findings, or conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies, WCER, or cooperating institutions.

# Effects of Testing Accommodations on Math and Reading Scores: An Experimental Analysis of the Performance of Fourth- and Eighth-Grade Students With and Without Disabilities[1]

**Ryan J. Kettler, Bradley C. Niebling, Andrew A. Mroch, Elizabeth S. Feldman, and Markeda L. Newell**

The 1997 reauthorization of the Individuals With Disabilities Education Act (IDEA, 1997) mandated that *all* students be included in large-scale school accountability systems to the extent possible. The passage of the No Child Left Behind Act (2002) further reinforced this position and increased the number of grade levels at which the achievement of all students must be measured. These policy shifts brought to the forefront the question of how to meaningfully include and accurately assess students with disabilities. Before the 1997 revision to IDEA, many schools excluded a large percentage of students with disabilities from large-scale achievement tests (Elliott, McKevitt, & Kettler, 2002). The last 5 years have seen a steady growth in the number of research studies focusing on testing accommodations, changes made to assessments so they are more valid for students with disabilities. The focus of most of the recent studies has been to determine whether testing accommodations have a greater impact on the scores of students with disabilities (SWD) than on the scores of students without disabilities (SWOD), a key characteristic of a valid testing accommodation (Phillips, 1994; Thompson, Blount, & Thurlow, 2002). Another body of recent research has looked at the alignment of testing accommodations with individual student needs (e.g., Elliott, Kratochwill, & McKevitt, 2001; McKevitt & Elliott, in press; Schulte, Elliott, & Kratochwill, 2001). This research has used single-case designs to evaluate testing accommodations. The current study utilized single-case methodology to determine what effects accommodations have on test results for individual students and groups of students (i.e., SWDs and SWODs).

### *Definition of Testing Accommodations*

The *Standards for Educational and Psychological Testing* (American Educational Research Association, 1999) defined testing accommodations as changes in the standard assessment process, including modifications to the test itself, made because an individual's disability requires changes for the test to be a valid measure. The inclusion of modifications to the test itself signified a departure from typical definitions, which have distinguished modifications from accommodations.

Testing accommodations are designed to remove barriers that would otherwise keep students with disabilities from demonstrating their true abilities on a test (Schulte et al., 2001). Elliott et al. (2002) invoked the analogy of eyeglasses used during a driving test to explain the notion of a testing accommodation. If eyeglasses were not allowed on a driving test, a perfectly capable driver with a visual impairment would probably fail, whereas a capable driver with no such impairment would pass the test. In the case of the good driver with no visual impairment, the driving test was valid. In the case of the person with the visual impairment, the barrier

[1] Paper presented at the annual convention of the National Association of School Psychologists, Toronto, Ontario, Canada, April 10, 2003.

created by that disability prevented the test from measuring the individual's driving ability. Accommodations, like the eyeglasses, are changes in the assessment procedure that should be helpful only to the individuals for whom they are designed.

### *Research on Testing Accommodations*

In theory, valid testing accommodations are useful only for the individual to whom they are tailored. They are useful for the person with the disability, but should be useless, perhaps even harmful, for a person without a disability. A second characteristic of testing accommodations, which differs from the first in subtle but important ways, is that they are individualized. Researchers of testing accommodations have tried to examine individualized accommodations that are differentially beneficial to students with disabilities.

***Differential boost***. One type of evidence that a testing accommodation helps provide a valid representation of the intended construct is a larger increase in the scores of SWDs than in those of SWODs, in moving from nonaccommodated to accommodated conditions. This "differential boost" (Phillips, 1994) is empirical evidence that the accommodations are helping to level the playing field and create assessments that work comparably for SWDs and SWODs. The notion of differential boost helps testing accommodation users and researchers navigate between two equally undesirable outcomes: tests that SWDs cannot meaningfully access, and inflated scores that are invalid indicators of SWDs' achievement.

Tindal and Fuchs (1999) considered differential boost in a large meta-analysis of research on testing accommodations. These researchers reviewed a large spectrum of accommodations via a framework suggested by the National Center on Educational Outcomes. They concluded that allowing students extended time typically helped SWDs more than SWODs. Tindal and Fuchs also found that the most effective accommodations were: (a) providing large print or Braille for the visually impaired, and (b) reading problems aloud for students with disabilities in math. Allowing the student to answer within the actual test booklet was not typically an effective accommodation, but allowing dictation to a scribe appeared helpful for some students. Finally, assistive devices such as calculators were found to be helpful in some cases, but detrimental in others.

***Individualized accommodations***. The Individuals With Disabilities Education Act (IDEA, 1997) mandates that testing accommodations be chosen on an individual basis. Individualized education plans (IEPs) must be designed for SWDs. These plans are drafted by education teams that include regular and special education teachers, parents, test interpreters, and resource administers. Because teachers are the individuals most likely to administer the accommodations, they are often influential in determining which accommodations are appropriate for each student. Although law and logic push teachers into the role of selecting individual accommodations, current research (Helwig & Tindal, 2003) indicates that teachers are not very accurate predictors of which students will benefit from testing accommodations or which testing accommodations will be helpful.

In two recent studies (Fuchs, Fuchs, Eaton, Hamlett, Binkley, & Crouch, 2000; Fuchs, Fuchs, Eaton, & Hamlett, 2000), Fuchs, Fuchs, and colleagues used a four-phase, data-driven method to individualize testing accommodations. In both studies, the researchers used shortened assessments to test SWDs and SWODs in one standard condition and in several accommodated conditions. In Phase I of the Fuchs, Fuchs, Eaton, Hamlett, Binkley, et al. (2000) study, the researchers found that a differential boost resulted from having students read aloud, but not from providing extended time or large print. In Phase I of the Fuchs, Fuchs, Eaton, Hamlett, and Karns (2000) study, the researchers found a differential boost on problem-solving measures, but not on computations or concepts and applications measures. During Phases II and III of both studies, participants were fitted with individualized accommodations packages. The researchers first used evidence from the Dynamic Assessment of Testing Accommodations (DATA) in Phase I to decide which students with disabilities should have accommodations, then had teachers identify which students with disabilities should have accommodations. In Phase IV, participants in both studies completed a full assessment in both accommodated and nonaccommodated conditions. In both cases, DATA improved upon teacher selection of accommodations.

In another line of research, Elliott and colleagues (Elliott et al., 2001; McKevitt & Elliott, in press; Schulte et al., 2001) have investigated individualized accommodations using a single-case design, requiring participants to take assessments in both accommodated and nonaccommodated conditions. In each of these studies, accommodation packages were selected using student IEPs and the Assessment Accommodations Checklist (AAC; Elliott, Kratochwill, & Schulte, 1999), an instrument designed to help individuals systematically select from the broad range of available testing accommodations. Elliott et al. (2001) tested SWDs and SWODs on a variety of math and science performance assessments, finding that the effect size for SWDs (.88) was double the effect size for SWODs (.44). On an individual level, accommodations had a medium to large effect on 78% of SWDs and 55% of SWODs. Schulte et al. (2001) used this same design with the TerraNova (CTB/McGraw-Hill, 1997) math test, finding that accommodations had a large differential effect on multiple-choice questions but not on constructed-response questions. McKevitt and Elliott (in press) compared the effects of teacher-recommended accommodations with and without a read-aloud accommodation, using this single-case design with the TerraNova reading test. They found that teacher-recommended accommodations from the AAC did not help either group, whereas a read-aloud accommodation combined with teacher accommodations boosted the scores of both SWDs and SWODs. The current study utilizes both the AAC to identify individualized packages of accommodations and a single-case design to examine the performance of SWDs and SWODs with and without testing accommodations on research versions of math and reading tests from a popular large-scale achievement test.

### *Summary*

Educators need valid testing accommodations to include many SWDs in large-scale assessment. Valid accommodations attempt to make the results for SWDs comparable to the results for SWODs on nonaccommodated assessments. Good assessment accommodations are individualized and assessed on an individual, and then aggregated, basis. Although many researchers (e.g., Thompson et al., 2002) have attempted to identify testing accommodations that meet these criteria, few have been able to do so. The current study is designed to address the

need for information about the effects of individualized testing accommodations on the test scores of students with and without disabilities.

## Research Question and Prediction

The current study focused on the following question: *What effects do accommodations have on test results for individual students and for groups of students (i.e., students with and without disabilities)?* To answer that question, we examined test score data from alternate forms of a research version of a widely used, standardized math and reading test administered under accommodated and nonaccommodated conditions. Intraindividual comparisons of each participant's scores via single-case analysis and interindividual comparisons of the scores of SWDs and SWODs provided the primary evidence of the effect of accommodations. We predicted that accommodations would significantly improve scores for SWDs but would have a negligible effect on the scores of SWODs. This prediction was based on theory and previous research (Fuchs, Fuchs, Eaton, Hamlett, Binkley, et al., 2000; Phillips, 1994; Schulte et al., 2001; Tindal & Fuchs, 1999) indicating that valid testing accommodations should improve the scores of SWDs without changing the scores of SWODs.

## Method

### *Participants*

The primary participants in the current study included 118 fourth-grade students and 78 eighth-grade students. Among the fourth-grade students, 49 met disability criteria in the state of Wisconsin. The fourth-grade sample included slightly more males ($n$ = 67) than females ($n$ = 51) and was primarily European-American ($n$ = 112). Among the eighth-grade students, 39 met disability criteria in the state of Wisconsin. This sample was also skewed toward males ($n$ = 52) and primarily European-American ($n$ = 76).

The secondary participants in this study were teachers ($n$ = 24). Each teacher was given a small stipend to help recruit student participants, coordinate data collection, and provide input on testing accommodations.

### *Materials*

***TerraNova***. We used four research editions of the TerraNova Multiple Assessment Battery (CTB/McGraw-Hill, 1997)—two math subtests (Form A and Form B) and two reading subtests (Form A and Form B)—to test participants' achievement levels. According to the teacher directions, the math tests were designed to measure broad mathematical ability while matching traditional curriculum. The tests begin with sections on computation and estimation, then move to a mathematical applications section that requires critical thinking and problem solving. The reading tests were designed to measure reading comprehension, language expression, vocabulary, and reference skills. Test questions, linked to overarching themes within each section, target students' levels of understanding, interpretation, and evaluation of concepts within each passage.

The TerraNova items were designed based on advice from panels of teachers, administrators, and content specialists. Item writers developed a large pool of developmentally and content-appropriate items, which were tested in a nationwide study in 1995. The developers then used item response theory, along with careful attention to possible bias based on ethnicity, race, gender, religion, or age, to select the final item pool.

The research editions of the TerraNova designed for this study were intended to be shorter forms of the test that would be relatively equivalent to the full forms. The fourth-grade versions had coefficient alpha reliabilities of .87 for both forms of the math test and .92 for both forms of the reading test. The eighth-grade versions had reliabilities of .86 on both math forms and .92 for both reading forms.

***Assessment Accommodations Checklist.*** The AAC (Elliott et al., 1999) is a guide used by educators to plan and document the use and effect of testing accommodations. Because it lists and categorizes the many types of accommodations that may be used in educational assessment, it can help teachers organize their thinking when planning accommodation packages. In the current study, the AAC provided a common framework for accurately defining the key independent variable of testing accommodations. It also provided a guide for enhancing treatment or intervention fidelity.

### *Procedures*

Students with disabilities were assigned accommodations based on their IEPs. Researchers then randomly selected an SWOD to pair up with each SWD. When testing under the accommodated condition, both students in the pair received the same set of accommodations. Project assistants (i.e., graduate students trained in administration of large-scale standardized tests) gave the students a general overview of the research before administering the four tests. Although the order of the conditions between pairs was randomly varied, each pair received one accommodated and one nonaccommodated math test and one accommodated and one nonaccommodated reading test. Within pairs, the students took the tests in the same order. Reliable scoring was achieved by having the CTB/McGraw-Hill staff score the test in the same fashion that they score large-scale assessments for states and large school districts.

### *Design and Analysis*

This study featured a 2 (Disability Status) × 2 (Testing Condition) × 2 (Grade) × 2 (Test Content Area) × 2 (Order) mixed design. The independent variables and their levels were disability status (i.e., SWDs and SWODs), accommodation condition (accommodated and nonaccommodated), grade (fourth and eighth), test content area (math and reading) and order (accommodated first or nonaccommodated first). The dependent variables were total reading test score and total math test score. To determine what effects accommodations had on test results for groups of students, we used a 2 × 2 × 2 analysis of variance (ANOVA) for each grade and test content area. Means, standard deviations, and effect sizes were computed to compare the effect of testing accommodations on each group. The following formula was used to compute effect sizes: (mean accommodated score - mean nonaccommodated score)/standard deviation of the nonaccommodated score). We also assessed test performance on an individual basis by counting

the number of SWDs and SWODs who moved either up or down one proficiency level according to criteria used for the Wisconsin Knowledge and Concepts Examination (Wisconsin Department of Public Instruction, 2003; see Table 1).

## Results

Data analysis revealed mixed results for our prediction that SWDs would show significant improvement for the accommodated test compared to the nonaccommodated test. Although fourth-grade students' reading results, on average, supported our prediction, their math results did not. Results from the group of eighth-grade students also failed to support our hypothesis that accommodations would significantly boost test scores. These results are examined in more detail next and highlighted in Tables 2–5.

### *Fourth Graders' Test Performance*

On the reading tests, fourth-grade SWDs were more affected than SWODs by the difference in accommodation condition. The interaction between disability status and accommodation condition was statistically significant, $F(1, 116) = 7.58$, $p < 0.05$. On the nonaccommodated reading test, fourth-grade SWDs averaged a standard score of 599.94, whereas SWODs averaged a standard score of 661.65 (see Table 2). In the accommodated condition, SWDs averaged a score of 618.92 on the reading test, whereas SWODs averaged 666.36. Although both groups of students improved, the improvement of SWDs had an effect size of .42, whereas the effect size of SWODs was .13. When individual performance was considered, 43% of SWDs moved up at least one math proficiency level, 51% stayed on the same level, and 6% moved down one level when they received accommodations (see Table 3). With regard to SWODs, 22% moved up at least one proficiency level, 68% stayed on the same level, and 10% moved down one level when they received accommodations.

On the math tests, fourth-grade SWDs were not affected differently than SWODs with regard to accommodation condition. The interaction between disability status and accommodation condition was not statistically significant, $F(1, 115) = 2.83$, $p > .05$. Among fourth-grade participants, SWDs averaged a standard score of 594.55 on the nonaccommodated math test, whereas SWODs averaged a standard score of 656.68 (see Table 2). In the accommodated condition, SWDs averaged a score of 615.76 on the math, whereas SWODs averaged 668.97. While both groups of students on average improved, the improvement of SWDs had an effect size of .46, whereas the effect size of SWODs was .27. For individual students, 41% of SWDs moved up at least one math proficiency level, 49% stayed on the same level, and 10% moved down at least one proficiency level when they received their individualized accommodations (see Table 3). For the SWODs, 29% moved up, 55% stayed on the same level, and 16% moved down at least one proficiency level when they received accommodations.

### *Eighth Graders' Test Performance*

On the reading tests, eighth-grade SWDs were not affected differently than SWODs by the accommodation condition. The interaction between disability status and accommodation condition was not statistically significant, $F(1, 75) = .01$, $p > .05$. On the nonaccommodated reading test, eighth-grade SWDs averaged a standard score of 640.47, whereas SWODs averaged a standard score of 695.00 (see Table 4). In the accommodated condition, SWDs averaged a score of 649.26 on the reading test, whereas SWODs averaged 703.64. The improvement of SWDs had an effect size of .25, compared to an effect size of .29 for SWODs. When performance of individual students was considered, 29% of SWDs moved up at least one reading proficiency level, 61% stayed on the same level, and 11% moved down one proficiency level due to accommodations (see Table 5). For SWODs, 26% moved up, 69% stayed on the same level, and 5% moved down at least one proficiency level when they received accommodations.

On the math tests, eighth-grade SWDs were not affected differently than SWODs by the accommodation condition. The interaction between disability status and accommodation condition was not statistically significant, $F(1, 76) = .10$, $p > .05$. For eighth-grade participants, SWDs averaged a standard score of 656.31 on the nonaccommodated math test, whereas SWODs averaged a standard score of 728.10 (see Table 4). In the accommodated condition, SWDs averaged a score of 662.92 on the math test, whereas SWODs averaged 732.41. The improvement of SWDs due to accommodations resulted in an effect size of .17, compared to the effect size of .13 for SWODs. The use of testing accommodations resulted in 18% of SWDs moving up at least one math proficiency level, while 69% stayed on the same level, and 13% moved down a level (see Table 5). For SWODs, 33% moved up at least one proficiency level, 51% stayed on the same level, and 15% moved down a level when they received an accommodation.

## Discussion

This study was influenced by the need for sound information on the effects and consequences of testing accommodations on fourth- and eighth-grade students' test scores. We examined changes in reading and math test scores for SWDs and SWODs in accommodated and nonaccommodated testing conditions. To understand this issue, we compared the scores of SWDs and SWODs. The results of our analyses were mixed. For fourth-grade students, testing accommodations provided a larger effect for SWDs than SWODs on both the mathematics and the reading tests. The effects of testing accommodations on both the mathematics and reading tests fell in the small to medium effect size range (Cohen, 1992) for SWDs, whereas the testing accommodations had only a small effect for SWODs. This finding is consistent with previous research. For example, Schulte et al. (2001) found small to medium effect sizes for fourth-grade SWDs on a standardized mathematics test, but only a small effect size for SWODs. Also supporting the differential benefit of testing accommodations for fourth-grade students was our finding that a higher percentage of SWDs moved up at least one proficiency level, due to testing accommodations, than did SWODs. Despite these indications of a differential boost in test performance for SWDs, it is important to note that the only statistically significant difference in test score change for the group was due to testing accommodations for the fourth-grade reading test. Moreover, for this study, it was permissible to have reading support on the reading test. This accommodation is controversial and typically considered to invalidate reading test scores on

most state assessments. Thus, users must take caution in interpreting the effects of testing accommodations across subject areas.

Our results for eighth-grade students were more mixed than the fourth-grade students' results. Among eighth-grade students, the effects of testing accommodations depended on test subject matter. The effects of testing accommodations on the mathematics tests were slightly higher for SWDs than for SWODs. Effect sizes for both groups, however, fell into the small effect range (Cohen, 1992). The findings on the reading tests were the opposite of those found on the mathematics tests. The effects of testing accommodations on the reading tests were slightly lower for SWDs than for SWODs. Again, effect sizes for both groups fell into the small effect range (Cohen, 1992). Although the small effect sizes of testing accommodations for SWODs are consistent with previous research (e.g., Schulte et al., 2001), the small effect sizes for SWDs are not. Results for eighth-grade students were also mixed on the individual level. Among eighth-grade students, a higher percentage of SWDs than SWODs moved up at least one proficiency level on the reading tests due to testing accommodations, whereas more SWODs than SWDs moved up at least one proficiency level on the mathematics tests. Furthermore, there was no disability status by testing condition interaction for the mathematics or reading tests. Stated another way, the score changes from nonaccommodated to accommodated testing conditions were not significantly different for SWDs than SWODs.

### *Interpretation and Generalization of Results*

Overall, it appears that testing accommodations may be beneficial for fourth-grade students, and they appear to provide the differential boost described by Phillips (1994) on a standardized reading test. Although some evidence does indicate this differential boost for SWDs is present on the fourth-grade mathematics test as well, the results are not as strong as those for the reading test. Furthermore, results from eighth-grade students indicate minimal positive effects for SWDs and SWODs, and no differential boost for either group. Another piece of evidence suggesting that testing accommodations may not have a meaningful impact on test scores for a large number of students was the impact of accommodations on movement between proficiency levels. Specifically, for both fourth and eighth graders, one half to three quarters of *all students* either remained within the same proficiency level or dropped at least one proficiency level when administered a test with accommodations. In fourth grade, it appears that the testing accommodations did remove some or all of the construct-irrelevant variance present in the tests for SWDs. However, given that SWODs also benefited somewhat from the testing accommodations in fourth grade, and all groups of eighth-grade students benefited from testing accommodations, testing accommodations may have been eliminating some construct-relevant variance.

Again, results for the fourth graders are consistent with the findings of Schulte et al. (2001). Although the effects of testing accommodations for the eighth graders in this study are small, the pattern of differential benefit for SWDs and SWODs is similar to that found in previous research. That is to say, there is evidence from previous research indicating a lack of a differential benefit of testing accommodations for SWDs and SWODs. For example, results from a study conducted by Elliott et al. (2001) indicate similar patterns of benefit of testing accommodations for SWDs and SWODs. This study, examining the effects of testing accommodations for SWDs and SWODs on science and mathematics performance assessments,

shows medium to large effect sizes for more than 75% of SWDs on accommodated testing conditions when compared to nonaccommodated testing conditions, indicating a generally positive effect on these students' test scores. There were two other interesting results from this study. First, almost 20% of SWDs showed a negative effect on accommodated testing conditions when compared to nonaccommodated testing conditions. This finding could indicate that the chosen accommodations packages that negatively affected performances simply were not good fits for the individual students. Second, approximately 50% of SWODs positively benefited from accommodations when compared to nonaccommodated testing conditions, although to a lesser degree than SWDs. This could be evidence that a number of students who approach, but do not meet, the criteria for learning disabilities could benefit from testing accommodations. It could also indicate that some accommodations used in this study functioned as bonuses rather than accommodations, increasing the scores of all students rather than decreasing construct-irrelevant variance.

Another study looking at the effects and consequences of testing accommodations examined performance on a standardized reading test (McKevitt & Elliott, 2001). Instead of administering one test with accommodations and another parallel form of the test without accommodations to SWDs and SWODs, the researchers split one standardized reading test in half and had students take half of the test with accommodations and half of the test without accommodations. Results indicated that receiving a package of accommodations was not differentially beneficial for SWDs or SWODs, although gains were found for individual students within each group.

This last finding from McKevitt and Elliott (2001) is key to understanding the results from the current study. Although the only significant differential boost for SWDs was found in fourth-grade reading, there were individual students in fourth and eighth grade who benefited from the testing accommodations on both the mathematics and the reading tests. Given the highly individualized selection of testing accommodations and the heterogeneity of SWDs' and SWODs' academic skills, it is not surprising that there was a wide variety of results due to accommodations. Recent research done by Helwig and Tindal (2003) indicated that teachers were no more successful in predicting which students would benefit from having a mathematics test read aloud to them than random chance would predict. Given the wide variety of the effects of testing accommodations for students in our study, it is possible that the teachers who recommended testing accommodations for the student participants had difficulty with the selection process as well. It should be noted, however, that the current study did not directly examine this particular phenomenon.

### *Limitations of the Study and Future Research*

One potential limitation of this study is that we examined only two test content areas and two grade levels. Specifically, we examined only mathematics and reading, although students are often tested in science and social studies as well. Nonetheless, the current study provides evidence that both subject matter and grade level have an impact on the effects of testing accommodations on test scores, results similar to those found in previous research (Trimble, 1998; Fuchs, Fuchs, Eaton, & Hamlett, 1999). The No Child Left Behind Act (2002) mandated the standardized testing of all third- through eighth-grade students. Given this mandate, an understanding of the impact of testing accommodations has become even more important than it

was previously. Future research on the effects of testing accommodations on student performance across multiple grades and subject areas will assist in this process.

A second and related potential limitation of the current study is its inability to explain why some students performed worse when using testing accommodations. Is this finding the result of a poor accommodations-to-student match? Were the accommodations inappropriate for these students, given the required tasks? Were the accommodations inappropriately implemented? These are questions that warrant further investigation.

A third potential limitation of the current study is its failure to operationalize the accommodations and their implementation. The resulting process, although more closely approximating the implementation of accommodations in practice, can lead to less than perfect experimental research. The AAC (Elliott et al., 1999) provided an excellent resource list from which to choose accommodations, but it did not provide definitions for each accommodation. Furthermore, the implementation of selected accommodations was not operationalized anywhere in this study's data collection procedures. For example, one of the most common accommodations offered to students in this study was the provision of extra time. Test administrators differed, however, in their interpretation of what constitutes extra time. Some thought the accommodation occurred whenever extra time was offered, whether it was actually used or not. Others considered the accommodation to have occurred only when the student took advantage of extra time to finish the test. Thus, the impact of a single accommodation—which again cannot be determined by the current study—could be different depending on how the accommodation was administered and used. Future researchers should define exactly how each of the available accommodations is to be administered, to eliminate operationalization as a source of potential variance.

## Conclusions

The results of this study indicate that testing accommodations can have a statistically significant and meaningful impact on the reading test scores of fourth-grade SWDs. Testing accommodations also appear to have a meaningful impact on mathematics test scores for fourth-grade SWDs, although the change was not statistically significant. For eighth-grade students, the effects of testing accommodations were small for both SWDs and SWODs. These results seem to support previous research indicating the generally positive, yet highly individualized, impact of testing accommodations on test scores.

The results of this study suggest fruitful directions for future research. More research is needed on the apparently highly individualized nature of the impact of testing accommodations. What are the student characteristics that influence the effects of testing accommodations? What factors influence educators' selection of testing accommodations for specific students? In what way do grade and subject matter interact with these factors to influence the effects of testing accommodations? Furthermore, the current study suggests that future research should more carefully operationalize the accommodation procedures actually used. Finally, continued use of single-case methods to examine the impact of single accommodations when compared to packaged accommodations will give a more accurate picture of the true impact of testing accommodations on test scores.

In the end, the ultimate goal of testing accommodations is to allow SWDs to fully participate in state accountability systems without providing them with an "unfair advantage" over their peers without identified disabilities. Based on the results of this study, it appears that testing accommodations can provide a differential boost in scores for SWDs over SWODs on some tests. These results provide evidence that accommodations may increase the validity of scores for some SWDs, while also suggesting that a number of students not identified as having a disability might benefit from accommodations. Further research is needed to find valid, individualized testing accommodations that result in comparable test scores for students with and without disabilities.

## References

American Educational Research Association. (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

CTB/McGraw-Hill. (1997). *TerraNova Multiple Assessments Battery.* Monterey, CA: Author.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

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*(1), 3–24.

Elliott, S. N., Kratochwill, T. R., & Schulte, A. G. (1999). *Assessment accommodations checklist.* Monterey, CA: CTB/McGraw-Hill.

Elliott, S. N., McKevitt, B. C., & Kettler, R. J. (2002). Testing accommodations research and decision making: The case of "good" scores being highly valued but difficult to achieve for all students. *Measurement and Evaluation in Counseling and Development, 35,* 153–166.

Fuchs, L. S., Fuchs, D., Eaton, S. B., & Hamlett, C. (1999, June). *Reading test accommodations for students with learning disabilities: Using the Dynamic Assessment of Test Accommodations (DATA) to supplement teacher judgments.* Paper presented at the Council of Chief State School Officers National Conference on Large-scale Assessment, Snowbird, UT.

Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children, 67*(2), 67–81.

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

Helwig, R., & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Council for Exceptional Children, 69*(2), 211–225.

Individuals With Disabilities Education Act Amendments of 1997, Pub. L. No. 105-17, 111 Stat. 37 (codified as amended at 20 U.S.C. § 1400 *et seq.*).

McKevitt, B. C., & Elliott, S. N. (in press). The use of testing accommodations on a standardized reading test: Effects on scores and attitudes about testing. *School Psychology Review.*

No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2002).

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

Schulte, A. G., Elliott, S. N., & Kratochwill, T. R. (2001). Experimental analysis of the effects of testing accommodations on students' standardized achievement test scores. *School Psychology Review, 30*(4), 527–547.

Thompson, S., Blount, A., & Thurlow, M. (2002). *A summary of research on the effects of test accommodations: 1999 through 2001* (Technical Report 34). Minneapolis: University of Minnesota, National Center of Educational Outcomes. Retrieved March 28, 2003, from http://education.umn.edu/NCEO/OnlinePubs/Technical34.htm

Tindal, G., & Fuchs, L. (1999). *A summary of research on testing accommodations: What we know so far.* Lexington: University of Kentucky Mid-South Regional Resource Center.

Trimble, S. (1998). *Performance trends and use of accommodations on a statewide assessment: SWDs in the KIRIS on-demand assessments from 1992–93 through 1995–96* (State Assessment Series, Maryland/Kentucky Report). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Wisconsin Department of Public Instruction. (2003). *Proficiency score standards 1997–98 through 2001–02: For the Wisconsin Student Assessment System (WSAS) Knowledge and Concepts Examinations.* Madison, WI: Author. Retrieved March 28, 2003, from http://www.dpi.state.wi.us/dpi/oea/profdesc_old.html

Table 1
*Wisconsin Knowledge and Concepts Examination Proficiency Levels*

| | Fourth grade | | Eighth grade | |
|---|---|---|---|---|
| Proficiency level | Reading | Math | Reading | Math |
| Advanced | 684+ | 659+ | 718+ | 750+ |
| Proficient | 625–683 | 623–658 | 672–717 | 718–749 |
| Basic | 600–624 | 581–622 | 655–671 | 674–717 |
| Minimal | 623 | 580 | 654 | 673 |

Table 2
*Fourth-Grade Means, Standard Deviations, and Effect Sizes by Disability Status and Test Content Area*

| Group | Nonaccommodated test | | Accommodated test | | Effect sizes |
|---|---|---|---|---|---|
| | M | SD | M | SD | |
| Reading | | | | | |
| SWD | 599.94 | 45.37 | 618.92 | 24.87 | .42 |
| SWOD | 661.65 | 37.17 | 666.36 | 33.88 | .13 |
| Math | | | | | |
| SWD | 594.55 | 46.45 | 615.76 | 40.74 | .46 |
| SWOD | 656.68 | 45.76 | 668.97 | 42.07 | .27 |

Table 3
*Number of Fourth-Grade Students at Each Proficiency Level Across Disability Status and Condition*

| Group | Students with disabilities | | Students without disabilities | |
|---|---|---|---|---|
| | Nonaccommodated | Accommodated | Nonaccommodated | Accommodated |
| Reading | | | | |
| Advanced | 1 | 0 | 20 | 25 |
| Proficient | 11 | 20 | 40 | 39 |
| Basic | 14 | 21 | 7 | 3 |
| Minimal | 23 | 8 | 2 | 2 |
| Math | | | | |
| Advanced | 3 | 7 | 30 | 31 |
| Proficient | 11 | 15 | 26 | 33 |
| Basic | 19 | 15 | 11 | 4 |
| Minimal | 16 | 12 | 2 | 1 |

Table 4
*Eighth-Grade Means, Standard Deviations, and Effect Sizes by Disability Status and Test Content Area*

| Group | Nonaccommodated test | | Accommodated test | | Effect sizes |
|---|---|---|---|---|---|
| | M | SD | M | SD | |
| Reading | | | | | |
| SWD | 640.47 | 35.19 | 649.26 | 30.67 | .25 |
| SWOD | 695.00 | 29.42 | 703.64 | 32.62 | .29 |
| Math | | | | | |
| SWD | 656.31 | 38.50 | 662.92 | 32.08 | .17 |
| SWOD | 728.10 | 34.44 | 732.41 | 33.56 | .13 |

Table 5
*Number of Eighth-Grade Students at Each Proficiency Level Across Disability Status and Condition*

| Group | Students with disabilities | | Students without disabilities | |
|---|---|---|---|---|
| | Nonaccommodated | Accommodated | Nonaccommodated | Accommodated |
| Reading | | | | |
| Advanced | 0 | 0 | 5 | 10 |
| Proficient | 6 | 10 | 28 | 23 |
| Basic | 10 | 8 | 3 | 5 |
| Minimal | 22 | 20 | 3 | 1 |
| Math | | | | |
| Advanced | 0 | 0 | 7 | 12 |
| Proficient | 0 | 2 | 16 | 13 |
| Basic | 14 | 13 | 15 | 14 |
| Minimal | 25 | 24 | 1 | 0 |