DOCUMENT RESUME

ED 423 632 EC 306 502

AUTHOR	Tindal, Gerald; Heath, Bill; Hollenbeck, Keith; Almond, Patricia; Harniss, Mark
TITLE	Accommodating Students with Disabilities on Large-Scale Tests: An Empirical Study of Student Response and Test Administration Demands.
PUB DATE	1998-00-00
NOTE	26p.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Disabilities; Educational Assessment; Elementary Education; Evaluation Methods; Grade 4; *Student Evaluation; *Test Results; *Testing; Testing Problems; *Verbal Communication
IDENTIFIERS	*Testing Accommodations (Disabilities)

ABSTRACT

In this study, fourth-grade special education students (n=78) and general education students (n=403) took a large-scale statewide test using standard test administration procedure and two major accommodations addressing response conditions and test administration. On both reading and math tests, students bubbled in answers on a separate sheet (the standard condition) for half the test and marked the test booklet directly (the accommodated condition) for the other half of the test. For a subgroup of students, the math test was read to them by a trained teacher. On the reading tests, general education students performed significantly higher than special education students. Performance, however, was not influenced by the response conditions and remained comparable whether students were required to bubble the answer sheet or allowed to mark the test booklet. The same findings occurred on the math tests with general education students performing significantly better than special education students and student performance not affected by response conditions. When the math test was orally read to students, general education students outperformed special education students; however, students in special education performed significantly higher when the math test was read by teachers than when they read the test themselves. (Contains 21 references and 4 tables.) (CR)

Running Head: ACCOMMODATING STUDENTS WITH DISABILITIES ON LARGE-SCALE TESTS

Accommodating Students with Disabilities on Large-Scale Tests: An Empirical Study of Student Response and Test Administration Demands

Gerald Tindal

Bill Heath

Keith Hollenbeck

University of Oregon

Patricia Almond

Oregon Department of Education

Mark Harniss

Northern Illinois University

Abstract

In this study, fourth-grade special and general education students took a large-scale statewide test using standard test administration procedures and two major accommodations addressing response conditions and test administration. On both reading and math tests, students bubbled in answers on a separate sheet (the standard condition) for half the test and marked the test booklet directly (the accommodated condition) for the other half of the test. For a subgroup of students, the math test was read to them by a trained teacher. Although no differences were found in the response conditions, an interaction was found in the test administration conditions (orally reading the test), supporting this accommodation for students with disabilities.

Accommodating Students with Disabilities on Large-Scale Tests:

An Empirical Study of Student Response and Test Administration Demands

With the most recent reauthorization of the Individuals with Disabilities Education Act (IDEA), students with disabilities must, to the greatest extent possible, be included in all large-scale, statewide testing programs. Generally, a multiple-choice test format is used in most of these assessment programs (Bond, Braskamp, & Roeber, 1996), in which teachers are presented booklets of test items that have been field-tested and an administration booklet detailing both the general conditions for giving the test and the specific verbatim directions to use during the administration. When the test administration is standardized, student scores are assumed to be comparable and the inferences made from student performance are, therefore, assumed to be more equitable: No student has an unfair advantage or disadvantage.

Although the use of standard administration conditions allows comparability across students, the validity of the inferences made on the basis of the outcomes (Messick, 1989) may be suspect if unrelated access skills needed to take the test actually impede performance. For example, students with reading problems may perform poorly on math tests, not because of their lack of mathematics proficiency, but because the test requires them to read a considerable amount of text: Many math test items contain extensive text describing a problem followed by more text providing multiple choices, all of which have to be read before the student can select the correct option. Low performance could be as much a function of poor reading skills as limited math proficiencies, restricting the inferences that can be made. Particularly with high-stakes decisions, such invalid inferences cannot be tolerated. For example, fully one third (17) of the 45 states using large-scale assessments require students to pass a statewide test for promotion or high school graduation (Bond, Braskamp, & Roeber, 1996). At the same time, the decision to make an accommodation (such as reading a math test), though widely adopted across many state practices (Siskind, 1993), frequently is not based on empirical data. Rather, "to avoid litigation when in doubt, the test administrator may want to err on the side of granting the required accommodation whenever feasible" (Phillips, 1994, p. 104). In conclusion, we are making important decisions using tests which require complex clusters of skills to complete, and for which accommodations frequently are allowed, all done in the absence of data.

The purpose of our research is to determine if two specific test accommodations (a) help students complete large-scale tests in a fair and equitable manner and increase the validity of inferences made from their performance, and at the same time (b) don't change the construct of what is being measured (in this study, reading and math) (Thurlow, Scott, & Ysseldyke, 1995). The accommodations investigated in this study are derived from a list of four general classes of modifications assembled by Thurlow, Scott, & Ysseldyke (1995): (a) timing and scheduling of the test, (b) setting in which the test is taken, (c) response demanded to complete the test (such as modifications in the test format or the use of assistive devices), and (d) presentation of the test to students (such as modifications to the test directions and the use of assistive devices or support modifications). In this study, we studied both a *response* (marking format) and a *presentation* (administration directions) accommodation.

The most extensive studies of test accommodations have been done with Educational Testing Services (ETS) on the Graduate Record Examination (GRE) and the Scholastic Aptitude Test (SAT) (Willingham et al., 1988). In general, they found that, between the standard and nonstandard administrations, there was (a) comparable reliability (Bennett, Rock, & Jirele, 1986; Bennett, Rock, & Kaplan, 1985, 1987); (b) similar factor structures (Rock, Bennett, & Kaplan, 1987); (c) similar item difficulties for disabled and nondisabled examinees (Bennett, Rock, & Kaplan, 1985, 1987); (d) noncomparable predictions of academic performance (with the nonstandard test scores less valid and SAT test scores substantially underpredicting college grades for students with hearing impairments) (Braun, Ragosta, & Kaplan, 1986); and (e) comparable admissions decisions (Benderson, 1988).

In an analysis of test content, Willingham et al. (1988) found that, although students with disabilities perceived the test to be harder, their performance was comparable to peers without disabilities. He also found that college performance was overpredicted when extended time was allowed.

In the end, these researchers recommend that those using any test results "(a) use multiple criteria to predict academic performance of disabled students, (b) give less weight to traditional predictors and more consideration to students' background and nonscholastic achievement, (c) avoid score composites, (d) avoid the erroneous belief that nonstandard scores are systematically either inflated or deflated, and (e) where feasible and appropriate, report scores in the same manner as those obtained from standard administrations" (ETS, 1990, Executive Summary Report).

The ETS research, however, is limited to college admission testing, all of which represents a limited group of tests for students with disabilities (e.g., college-bound secondary students). The number of students with disabilities who participate in such tests is very small (proportionately) and may not be representative of the larger group of such individuals (within any disability group or even in the general population).

Another small body of literature exists from the mid-1980s in which test accommodations are either proposed or investigated for students with disabilities. Some of this literature presents modifications and accommodations which sound sensible but have no empirical basis for adoption (Harrington & Morrison, 1981; Salend & Salend, 1985; Wood & Aldridge, 1985). Furthermore, some of the outcomes represent survey data and fail to report performance outcomes in relationship to modifications (McKinney, 1983), making judgments of validity difficult.

Nevertheless, two teams of researchers have compiled four studies in which test accommodations were empirically investigated. In a study by Grise, Beattie, and Algozzine (1982), about 350 students in fifth grade took the Florida State Student Assessment Test with seven different changes made in the format of the test. They found that students with learning disabilities performed slightly higher on the regular print version (vs. an enlarged version) on only one of six subsections. They also found 20% to 30% more students who were administered the modified version (vs. the regular print version) performed at mastery levels in various subsections of the test. In a comparable study using the same modifications with a third-grade sample of students (n = 345), Beattie, Grise, and Algozzine (1983) again found few differences on most subsections when comparing performance on the regular print version versus an enlarged print version. And, as in the other study, more students with learning disabilities mastered most of the skills when taking the modified test; on many skills, 20% more students reached mastery levels when the modified version was used than when the test was taken under standard conditions.

Tolfa-Veit and Scruggs (1986) conducted an empirical investigation focused on the use of separate answer sheets with 101 students in Grade 4 (19 students with learning disabilities). Although they found significant differences between general and special education students in the total number of items copied onto an answer sheet (97 versus 86, respectively), they found no significant differences in the percentage of items marked correctly (both groups were about 97% correct). Finally, in a study with 85 students with learning and behavioral disabilities, Scruggs, Mastropieri, and Tolfa-Veit (1986) coached students in several specific test-taking strategies. They found significant differences between the trained and no-treatment control students in word study and math concepts, although no significant differences were found on reading comprehension and math story problems.

In summary, the literature on test modifications is thin. The most significant problem is the lack of appropriate experimental and control groups and conditions. For the two studies from Florida (Beattie, Grise, & Algozzine, 1983; Grise, Beattie, & Algozzine, 1982), no general education students received the modified tests. For the answer sheet study (Tolfa-Veit & Scruggs, 1986), the task fails to appropriately reflect the complex demands of actual test conditions in which students must read the problem, solve it, and then fill in an answer sheet. Finally, in the last study (Scruggs, Mastropieri, & Tolfa-Veit, 1986), no general education students were included in the sample (in either the trained or no-treatment control groups).

In contrast, our study adds to this line of research by implementing a test accommodation with students in both special and general education. As Phillips (1994) has noted, (a) students with disabilities should take the standard administration if at all possible and (b) any accommodations from these standard testing conditions should be of little benefit to examinees with no such disabilities. These two features make the research design an important component of any study on accommodations because an interaction is being hypothesized over any main effects: To validate an accommodation, it must not only work with the targeted subgroup (e.g., students in special education) but also must not work for students in general education.

We not only endorse this logic but also believe the argument actually needs to be even more specific. Students with Individualized Educational Plans (IEPs) in reading and/or math can be assumed to have in common the need for an accommodation which neutralizes any access skills required to complete a math test which are unrelated to the skill being tested. The manner in which eligibility is conferred (that is, whether or not the student is in special education) or the etiology-type of disability (that is, whether the student has a designation of learning disabilities, speech-language, or behavioral disorders) is less relevant than the relationship between the need of the student as documented by the IEP and the demand of the test. Furthermore, for students with no such need (e.g., no IEP and therefore presumably not in special education and with no disability designation), the demands of the test should be irrelevant. However, even including this group does not provide a sufficiently strong test of the effect of an accommodation. Rather, to provide the most convincing empirical support for an accommodation, students with a specific need have to be compared to others without such a need who are otherwise comparable in achievement. With these issues in mind, we asked the following questions:

1. What is the effect on math and reading performance when students are allowed to mark their answer in the test booklet over that attained when students are required to bubble in an answer sheet? Is this effect similar for students in special and general education?

2. What is the effect on math performance when students have the math test read aloud to them? Is this effect similar for special and general education students? Is the effect similar when the accommodation is made for students likely to benefit from it—students with IEPs in reading/math versus those perceived to be low achieving in general education.

Methods

The study was conducted in 22 fourth-grade classrooms distributed across seven elementary schools. Testing at all schools occurred during the last part of May 1996. Teachers from both general and special education participated; 13 were female and 9 were male. All of them had elementary teaching certificates and 9 possessed master's degrees. The mean total years teaching was 16 (with a range of 2 to 33 years and a standard deviation of 10 years). The mean years teaching at fourth grade was 8 (with a range of 1 to 28 years and a standard deviation of 10 years).

Subjects

A total of 481 students participated, with the seven schools contributing from a low of 54 students to a high of 79 students (representing from 11% to 16% of the study participant population, respectively). Student age could be calculated for 463 students and ranged from just younger than 9 years to just older than 12 years, with 10.3 years the average. Female students totaled 228 students (48%), and male students totaled 251 (52%), with 2 missing records. Most of the students were White, with the largest minority group being Hispanic (16 or 3.5%), followed by American Indian (7 or 1.5%), Asian Pacific (4 or .9%) and Black (4 or .9%). For 409 students who completed the demographic information on the test form, about 75% (306 of 409) indicated that they had been in that school the previous year. When asked about their primary language on the test form, 374 students answered, with the greatest percentage indicating English as their first language (369, representing 97%)

and only 6 indicating English as a second language; 5 responded that they were Limited English Proficient (LEP).

Our analysis of students' educational status revealed 403 from general education (84%) and 78 from special education (16%). The students in special education were receiving assistance through 171 different Individualized Educational Plans (IEPs). For the 44 students with IEPs in reading, concurrent IEPs appeared for written expression (28), math (17), speech-language (15), language arts (13), spelling (13), study skills (1), behavior (1), and language (1). For the 20 students with IEPs in math, concurrent IEPs appeared in reading (17), written expression (14), speech-language (11), language arts (10), spelling (9), behavior (1), and language (1).

At the beginning of the study, teachers were asked to rank students from low (1) to high (n in the class) on achievement so that we could eventually compare students in special education (with IEPs in reading and/or math) with a subset of students who had been ranked as the lowest 5 and lowest 10 general education students on achievement in the class. Three weeks later, four teachers were asked to make the same ranking; all four teachers were very stable in ranking students in their classes on overall achievement. When we compared students with IEPs in reading and/or math to students in general education with the lowest 5 or 10 rankings on achievement, we obtained very comparable population proportions to the total group on the demographic characteristics.

Test Administration

During testing, student attendance was high. In reading, 95% of the students took part in both portions of the testing and 93% of the students attended both parts of the math testing. The reading data files were complete for 229 general education and 36 special education students in the two response conditions; in math, 198 general education and 38 special education students participated in the two administration conditions with complete data sets.

All students participated in the study on response accommodation by bubbling in an answer sheet and marking the booklet. The response accommodation study employed a design in which students were crossed with the accommodation and thus participated in both conditions. The State Department of Education split the reading and math tests into two booklets, with problems completed either by bubbling an answer sheet or directly marking the booklet; the order of administration was counterbalanced across the student population.

The presentation accommodation was investigated only with the math test and only with a subgroup of students. This part of the study employed a design in which students were nested within accommodation, with students randomly assigned to either one of the two conditions: Some students silently read the test while others (in different classrooms) listened as the teacher orally read the test. The read-aloud condition consisted of a math test being read in its entirety, including the general directions (for filling out the forms and taking the test), each specific problem, and all item choices for multiple choice problems. The reading of math problems was standardized to (a) prevent auditory cueing of correct options, (b) present reading assistance that was consistent with the problem type, and (c) avoid fast pacing of students in completing problems. All problems were read twice with students told to answer only after the problem was read the second time. An overhead of each page of the test booklet was prepared so the teacher could visually track students by pointing to the words/lines as the problems and choices were read. In all schools in which the math problems and multiple choices were read, graduate student proctors from a nearby university were utilized to ensure fidelity of treatment.

After testing was completed, the same graduate proctors transferred all answers from the booklets onto the standard bubble sheet. To establish reliability, 214 (45%) of the booklet-to-answer sheets were randomly chosen and checked. Exact matches (a correct transfer) were scored as 1 point, incorrect matches (an error in transfer) were scored as 0 points. Reliability for booklet-to-answer sheet transfer was .998 for reading first half, .999

for reading second half, 1.0 for math first half, and .984 for math second half. Once all the answer sheets were complete, items were hand scored as correct (1 point) or incorrect (0 points). Again, 45% of the student answer sheets were rescored to compute reliability; we attained coefficients of .999 for both reading and math. Finally, student answers for each problem were entered into a data file. Reliability also was analyzed for computer entry accuracy, which was perfect.

Data Analysis

After all data entry was completed and checked, student scores were statistically analyzed using a one-between (student status), one-within (response format), repeated measures analysis of variance for the response accommodation and a simple two-way analysis of variance for the presentation accommodation (student status and presentation accommodation, averaging over the two response format scores). Because of the large differences in sample sizes, students in general education were randomly divided into five groups and then these groups randomly sampled to conduct various comparisons with special education. First, the effect of bubbling in the answer sheet or marking the test booklet was analyzed for reading and then for math. In both of these analyses, a random group of general education students' performance was compared with special education students. Second, for the students who participated in the oral reading of the math test, the effect of the response accommodation (mark test booklet versus bubble answer sheet) was analyzed, first with a random sample of general education students and all special education students and subsequently with only low achievement-ranked general education students versus those in special education with IEPs in reading and/or math. Third, overall main effects for the administration condition of teacher versus student reading and the status of the student (general versus special education) were studied, and an interaction analysis was done for three groups when the teacher orally read the test: comparing (a) general education students with all special education students, (b) lowest ranked 10 general education students with special education students with an IEP in reading and/or math, (c) lowest ranked 5 general education students with the same special ed/IEP group. Follow-up contrasts have been calculated to ascertain simple effects within groups.

Results

On the reading test, a significant difference appeared between groups of students: General education students performed significantly higher than special education students, with $F(1, 131) = 68.4$, $p < .0001$). Performance, however, was <u>not</u> influenced by the response conditions and remained comparable whether students were required to bubble the answer sheet or allowed to mark the test booklet, with $F(1, 131) = .483$, $p = .4884$. No interaction was found between the status of the students and the response conditions, with $F(1, 131) = .047$, $p = .8282$. See Table 1. The same findings occurred on the math test. General education students performed significantly better than special education students, with $F(1, 131) = 34.815$, $p < .0001$. Performance was not affected, however, by the response conditions and students from general and special education performed equally well whether they bubbled in the answer sheet or marked their answers in the test booklet, with $F(1, 131) = .142$, $p = .7073$. Again, no interaction between student status and response condition was found, with $F(1, 131) = .163$, $p = .6868$. See Table 2.

When the math test was orally read to students, general education students outperformed special education students; however, performance was <u>not</u> influenced by the response conditions of bubbling the answer sheet or marking the test booklet. While a random sample of general education students performed significantly higher than special education students with $F(1, 40) = 19.700$, $p < .0001$, this performance was the same in either response condition, again reflecting no main effect for the response accommodation, with $F(1, 40) = .008$, $p = .9297$ or the interaction of student status with response condition ($F(1, 40) = 1.849$, $p = .1815$. See Table 3.

An analysis of the interaction between the administration of the math test (student silent reading or teacher oral reading) and the status of the student (general versus special education) was conducted to determine if the administration accommodation was uniform or differential in its effect on performance. In Table 4, we have reported the results for three different populations, comparing the lowest 10 ranked students in general education versus students in special education with IEPs in reading and/or math. For this analysis, the main effect was significant between between students' status, with $F(1, 160) = 32.730 < .0001$, as well as between student-teacher reading, with $F(1, 160) = 3.797$, $p = .0531$. In addition, their interaction was significant, rendering the two main effects for each of these factors not meaningful. Students in special education with IEPs in reading and/or math performed significantly higher when the math test was read by teachers than when they read the test themselves. In contrast, the performance of the 10 lowest achievement-ranked students in general education revealed no such improvements when teachers orally read the math test over that achieved when students silently read the math test), with $F(1, 160) = 9.049$, $p = .0031$. In the follow-up contrasts, no significant differences between the administration conditions were found for students in general education, although the differences were significant for students in special education. See Table 4.

Discussion

In the response accommodation for both a reading and a math test, we allowed students to mark their answers in the test booklet and compared their performance to the levels achieved when they took the test in the standard manner (bubbling in an answer sheet). We found no differences.

In the presentation accommodation, we had trained teachers to read the math test orally (the entire problem as well as all items on the multiple-choice test). We then compared outcomes for various groups of students, not only looking broadly at general versus special education students but also sampling the lowest ranked 5 or 10 students in the general education classroom and sampling students in special education with IEPs in the target area being tested. These various sampling plans allowed us both to focus the question on a critical sample and to provide a more balanced comparison with approximately equal sample sizes. The results were significant when reading was removed as a requisite access skill. This finding, however, needed to be qualified by the characteristics of the students, for not all of them were equally affected by the accommodation. When we defined an accommodation in relation to a common need for assistance via reading and/or math IEPs, it appeared that more valid inferences of math proficiency were possible when students had the test read to them.

Limitations

Our findings represent initial research to appear on the investigation of test accommodations, and our findings need to be interpreted within the context of the design we employed. For example, the response accommodation of marking the booklet was not generally any more effective than bubbling in an answer sheet. As a group, students performed at similar levels in both conditions; however, individuals within the two responses may have had higher scores when marking the booklet but the effect was removed when averaged with other students. Clearly, all group design studies suffer from this limitation. No absolute statements about the accommodation can be made for all students. Rather, in general and on the average, performance appears not to be affected by this accommodation.

In like manner, the presentation accommodation results are initial findings that need to be replicated with different subjects and using different designs. For example, we employed a design in which students were nested within the accommodation and randomly assigned to either the standard administration (student silently reads the test) or to the accommodated administration (teachers reads the test). It may be less confounding to use a design in which students are crossed with the accommodation and receive both of them (assuming they are counterbalanced in the order in which they are given). Although we believe that the subjects were comparable (matched by area of assistance and ranked as the lowest in achievement in the class), it is possible, though not likely, that slight between-group differences account for the findings.

And of course, future research needs to be done with older and other students, irrespective of the two particular designs we employed. For example, our study was conducted in fourth grade, when reading is just beginning to be used as an access skill to other content areas. And the math tests themselves, as well as the mathematics curriculum which they purport to reflect, may influence the degree to which reading is an important access skill. In later grades when mathematics algorithms become more complex and formula-specific, reading may become less important.

Interpretation

Our findings need to be interpreted in relation to both practice and measurement theory. For example, since no significant differences were found whether students marked the test booklet or bubbled in the answer sheet, teachers can make this accommodation decision on an individual basis without affecting the validity of any inferences made from test results. In our own discussions with teachers, we frequently hear how many students get confused in keeping track of the answer sheet and that once they are off in aligning the test problem number with the bubble number, all remaining problems become essentially random responses with the probability of being correct equal to chance. With this accommodation, teachers may let students simply focus on the problems, mark the test booklet directly, and then transcribe the items onto the bubble sheet.

While this accommodation appears easy, two issues should be considered, however, before immediately adopting it on a large-scale basis. Obviously, the process of transcribing students' answers from the test booklet to an answer sheet is time-consuming. Although university students and clerical staff were hired to complete this activity in our study, it is unlikely that schools have adequate personnel to do this for a great number of students. The other issue is the problem of reliability in transcribing student answers. We took great care and checked many of the protocols twice. And, although we were very reliable, the process is tedious and personnel may easily begin to drift. Valid inferences cannot be made from unreliable measurement.

Practice and measurement issues also arise when interpreting the findings from the presentation accommodation. As Phillips (1994) notes, if an accommodation is to be effective without changing the construct that is being measured, then perforce, an interaction must be obtained. If the accommodation is equally effective for all students, whether in general or special education, then two problems ensue. Practically speaking, the results simply raise the playing field, leaving students with disabilities the same relative distance as achieved without the accommodation. From a measurement perspective, when no interaction is present, it is likely that the construct is changed. In our finding, if the read-aloud is equally effective with all students, we need to view the test differently than if it is administered under standard conditions. Under these circumstances, students with a common need apparently are as equally affected as those without that need. We then have a situation in which we have changed the construct and the validity of the inferences is not enhanced (Thurlow, Scott, &Ysseldyke, 1995).

From an empirical point of view, it is difficult to place our findings in relationship to the work of Grise et al. (1982) and Beattie et al. (1983). Although they report significant effects from an accumulation of several minor accommodations to the test format, we find no effects from a singular accommodation in student responding. Likewise, although the outcomes from Tolfa-Veit and Scruggs (1986) seem to implicate bubbling in answer sheets as an inhibitor of performance for students with disabilities, we find no differences in outcomes when students mark the test booklet rather than shade in bubbles. It is very likely that the tasks in their study are not cognitively comparable to the real tasks presented within a large-scale, statewide test. Finally, the coaching study done by Scruggs et al. (1986) is radically different in treatment; our accommodations are far more limited to the test administration rather than strategies for taking tests. Therefore, it is difficult to determine how our findings relate to theirs. It may be that the read-aloud condition for students with disabilities is sufficient enough to remove the need for such intensive interventions. At the very least, it is unlikely that states are about to adopt coaching strategies as part of their test administration, although a read-aloud condition may be easier to both adopt and implement in a standard manner.

In summary, as states move into large-scale testing that includes students with disabilities, it is important to make appropriate accommodations. On the one hand, many accommodations may be useful for specific students and do not change the outcomes (such as marking the booklet instead of bubbling the answer sheet). And although we found no effect from using it, for some students this accommodation may be helpful. On the other hand, some accommodations change the outcomes, but differentially so (such as the teacher reading aloud the math problems and choices for students with reading/math IEPs). Student performance appears to be impeded by not using the accommodation and invalid inferences are being made when only the standard testing conditions are followed (Messick, 1989).

Implications for Practice

Increasingly, teachers need to consider accommodations in the manner in which tests are given and taken, because more states are relying on large-scale assessments, because of the new mandates of IDEA, and because the stakes in many of the decisions being made from these tests are indeed quite serious. Of course, it would be ideal if teachers could simply turn to the research and find a list of preferred and best practices in testing students with disabilities. Given the lack of empirical data now and in the near future, and given the less than uniform outcomes which are likely to eventually ensue, teachers can at least use the current study to develop a systematic decision-making process for determining which accommodations to consider.

First, any accommodation is likely to be listed as acceptable or unacceptable to use within the states guidelines. And although this list of accommodations may not be sacrosanct but simply reflects the conventional wisdom of individuals at either the state or local educational agency, teachers can begin to be sensitive to the decision-making process both by being aware of the accommodations and by knowing their implications. Two important considerations are the decision being made and the degree to which test results will be used to award or sanction individual students and teachers. If test data are being used to make high-stakes decisions, it may be critical to heed the advice of Phillips (1994) in using certain accommodations even though it is uncertain whether they change the construct being measured or provide a perceived unfair advantage. The result is likely to be more false positive decisions (e.g., awarding a Certificate of Initial Mastery in our study), which may have fewer negative effects in the end than making false negative decisions (e.g., denying students the CIM).

Second, as noted by the earlier research completed with the ETS group, decisions need to be made using multiple sources of information. In this study, we investigated the outcomes and impact of two types of accommodations in relation to perceived achievement and IEP assistance. Members of IEP meetings should consider a range of information when deciding to use an accommodation. For example, is it likely that the student needs such an accommodation? Has the student received this accommodation in the past? What is the likely effect of the accommodation with other, similar students? Answers to such questions may help prevent later difficulties from occurring, such as irate parents of a student without disabilities demanding similar accommodations for their child or parents of students with disabilities asking for blanket accommodations as a function of a disability designation rather than on the basis of need.

Finally, systematic data can be collected in the context of action research to begin justifying many of the decisions being made. Although such a strategy may not result in scientific research with threats to validity well controlled, it would certainly represent an improvement in the current decision-making process. For example, within the instructional program teachers could begin evaluating whether a student performs better with and without an accommodation using a single-subject design, in which the accommodation is alternately implemented and removed (withdrawal, A-B-A-B). Or a small group of students needing comparable areas of assistance could have an accommodation implemented in a lagged fashion (multiple baseline across subjects). Assuming comparability of measurement in the various phases and across the various subjects, such outcomes could represent a step forward in both ensuring that the accommodation is listed in the IEP and having some evaluative data supporting the accommodation.

References

Beattie, S., Grise, P., & Algozzine, B. (1993). Effects of test modifications on the minimum competency performance of learning disabled students. Learning Disability Quarterly, 6, 75–77.

Benderson, A. (Ed.) (1988). Testing, equality, and handicapped people [ETS Publication; Princeton, NJ]. Focus, 21, 23.

Bennett, R. E., Rock, D. A., & Jirele, T. (1986, February). The psychometric characteristics of the GRE General Test for three handicapped groups (ETS Research Report RR–86–6). Princeton, NJ: Educational Testing Service.

Bennett, R. E., Rock, D. A., & Kaplan, B. A. (1985, November). The psychometric characteristics of the SAT for nine handicapped groups (ETS Research Report RR–85–49). Princeton, NJ: Educational Testing Service.

Bennett, R. E., Rock, D. A., & Kaplan, B. A. (1987). SAT differential item performance for nine handicapped groups. Journal of Educational Measurement, 24 (1), 44–55.

Bond, L. A., Breskamp, D., & Roeber, E. (1996). The status report of the assessment programs in the United States: State Students Assessment Programs Database School Year 1994–1995. Oak Brook, IL: North Central Regional Educational Laboratory.

Braun, H., Ragosta, M., & Kaplan, B. (1986, October). The predictive validity of the Scholastic Aptitude Test for disabled students (ETS Research Report RR-86-38). Princeton, NJ: Educational Testing Service.

Educational Testing Service (1990). Testing persons with disabilities: A report for ETS programs and their constituents [Executive Summary Report]. Princeton, NJ: Author.

Grise, P., Beattie, S., & Algozzine, B. (1982). Assessment of minimum competency in fifth grade learning disabled students: Test modifications make a difference. Journal of Educational Research, 76 (1), 35–40.

Harrington, A. M., & Morrison, R. A. (1981). Modifying classroom exams for secondary LD students. Academic Therapy, 16 (5), 571–577.

McKinney, J. D. (1983). Performance of handicapped students on the North Carolina minimum competency test. Exceptional Children, 49 (6), 547–550.

Messick, S. (1989). Validity. In R. L. Linn (Ed.) Educational Measurement—Third Edition (pp. 13-104) New York: Macmillan.

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

Rock, D. A., Bennett, R. E., & Kaplan, B. A. (1987). Internal construct validity of a college admissions test across handicapped and nonhandicapped groups. Educational and Psychological Measurement, 47, 193–205.

Salend, S., & Salend, S. J. (1985). Adapting teacher-made tests for mainstreamed students. Journal of Learning Disabilities, 18 (6), 373–375.

Scruggs, T. E., Mastropieri, M., & Tolfa-Veit, D. (1986). The effects of coaching on the standardized test performance of learning disabled and behaviorally disordered students. Remedial and Special Education, 7 (5), 37–41.

Siskind, T. G. (1993). Modifications in statewide criterion-referenced testing programs to accommodate pupils with disabilities. Diagnostique, 18 (3), 233–249.

Thurlow, M. L., Scott, D. L., & Ysseldyke, J. E. (1995). A compilation of states' guidelines for accommodations in assessments for students with disabilities (Synthesis Report No. 18). Mpls., MN: University of Minnesota National Center on Educational Outcomes.

Tolfa-Veit, D., & Scruggs, T. E. (1986). Can learning disabled students effectively use separate answer sheets? Perceptual and Motor Skills, 63, 155–160.

Willingham, W. W., Ragosta, M., Bennett, R. E., Braun, D. A., Rock, D. A., & Powers, D. E. (Eds.). (1988). Testing handicapped people. Boston: Allyn & Bacon.

Wood, J. W., & Aldridge, J. T. (1985). Adapting tests for mainstreamed students. Academic Therapy, 20 (4), 419–426.

Table 1.

Reading Performance for Random Group of General vs. Special Education Students in Two Response Conditions (Mark Booklet or Bubble Answer Sheet)

	Count	Mean	Std. Dev.	Std. Err.
General Ed	136	17.5	4.0	.3
Special Ed	130	11.1	5.5	.5
Bubble Sht	133	14.511	5.7	.5
Mark Bk	133	14.278	5.9	.5
General Ed Bubble Sht	68	17.6	4.0	.5
General Ed Mark Bk	68	17.4	4.0	.5
Special Ed Bubble Sht	65	11.3	5.5	.7
Special Ed Mark Bk	65	11.0	5.7	.7

Table 2.

Math Performance for Random Group of General vs. All Special Education Students in Two Response Conditions (Mark Booklet or Bubble Answer Sheet)

	Count	Mean	Std. Dev.	Std. Err.
General Ed	120	22.0	3.6	.3
Special Ed	128	17.9	5.1	.5
Bubble Sht	124	19.9	4.6	.4
Mark Bk	124	19.8	5.2	.5
General Ed Bubble	60	22.0	3.5	.5
General Ed Mark Bk	60	22.0	3.8	.5
Special Ed Bubble Sht	64	18.0	4.7	.6
Special Ed Mark Bk	64	17.7	5.6	.7

Table 3.

Math Performance for Random Sample of General Education Students vs. Special Education Students in Oral Reading Administration Condition (Mark Booklet or Bubble Answer Sheet)

	Count	Mean	Std. Dev.	Std. Err.
General Ed	66	20.1	3.2	.4
Special Ed	18	15.5	3.6	.8
BubPerfM	42	19.2	3.7	.6
MrkPerfM	42	19.1	3.9	.6
General Ed Bubble Sht	33	20.0	3.3	.6
General Ed Mark Bk	33	20.3	3.1	.5
Special Ed Bubble	9	16.2	3.6	1.2
Special Ed Mark Bk	9	14.8	3.6	1.2

Table 4.

Interaction of Oral Reading of Math Test (Student or Teacher Reads) by Student Status (General vs. Special Education) for Three Student Sampling Plans

IEP Rdg-Mth/Rank <10[1]	Count	Mean	Std. Dev.	Std. Err.
General Ed	122	41.3	6.6	.6
Special Ed	42	33.5	8.3	1.3
Student Read	111	39.4	8.4	.8
Teacher Read	53	39.0	6.5	.9
Stdnt Reads-Gen Ed	89	41.6	7.1	.8
Stndt Reads-Spec Ed	22	30.5	7.7	1.6
Tchr Reads-Gen Ed	33	40.3	5.3	.9
Tchr Reads-Spec Ed	20	36.8	7.7	1.7

[1]Follow-up contrasts were not significant for general but are significant for special education

EC 306 502

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Accommodating Students with Disabilities on Large-Scale Tests: An Emperical Study of Student Response and Test Administration Demands

Author(s): Gerald Tindal and Others

Corporate Source:

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: Gerald Tindal SRM
Printed Name/Position/Title: Gerald Tindal / Professor
Organization/Address: Behavioral Research & Teaching
5262 University of Oregon
Eugene, OR 97403-5262
Telephone: 541/346-1640
FAX: 541/ 346-5689
E-Mail Address: geraldt@darkwing.uoregon.edu
Date: 03/26/99
geraldt@darkwing.uoregon.edu

(over)