DOCUMENT RESUME

ED 434 459 EC 307 457

TITLE	Test Changes: An Empirical Basis for Defining Accommodations. Project FORUM at NASDSE. A Synthesis Brief.
INSTITUTION	National Association of State Directors of Special Education, Alexandria, VA.
SPONS AGENCY	Special Education Programs (ED/OSERS), Washington, DC.
PUB DATE	1999-09-00
NOTE	18p.
CONTRACT	H159K70002
AVAILABLE FROM	Project FORUM; Tel: 703-519-3800 (Voice); Tel: 703-519-7008 (TDD); e-mail: matt@nasde.org; available in alternative formats.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Assistive Devices (for Disabled); *Disabilities; Elementary Secondary Education; *Performance Factors; Positive Reinforcement; *Standardized Tests; *Student Evaluation; Student Participation; Testing Problems; Time Factors (Learning)
IDENTIFIERS	Individuals with Disabilities Educ Act Amend 1997; *Testing Accommodations (Disabilities)

ABSTRACT

In light of the 1997 Amendments to the Individuals with Disabilities Education Act requiring the inclusion of students with disabilities in state and district assessments, this report summarizes findings of a research synthesis designed to provide empirically based data to inform states regarding decision making for assessment polices. The research reviewed a total of 114 studies. Results found: (1) extended time may be an appropriate accommodation; (2) positive effects for students with disabilities occurred as a result of test setting changes; (3) computer-based testing must be interpreted with caution due to rapidly changing technology and the lack of large-scale studies; (4) findings favor the use of familiar examiners, typically the student's classroom teacher, to provide accommodations and administer the test; (5) the use of large print or Braille and the reading aloud of math problems have been differentially effective; (6) dictation to a proctor or scribe appears to be effective in improving performance for students with and without disabilities; and (7) performance on standardized tests is greatly enhanced by reinforcement during the testing situation. An appendix includes the types of testing accommodations made for students with disabilities in the areas of setting, timing, response, presentation, test directions, and assistive devices. (CR)

NATIONAL ASSOCIATION OF STATE DIRECTORS OF SPECIAL EDUCATION, INC.

PROJECT FORUM

SYNTHESIS BRIEF

TEST CHANGES: AN EMPIRICAL BASIS FOR DEFINING ACCOMMODATIONS

SEPTEMBER 1999

SYNTHESIS BRIEF

Test Changes: An Empirical Basis for Defining Accommodations

PROJECT FORUM AT NASDSE SEPTEMBER 1999

Background and Purpose

A frequently voiced concern about implementing the 1997 Amendments to the Individuals with Disabilities Education Act (IDEA) is the requirement for inclusion of students with disabilities in state and district assessments in the absence of research findings for making decisions. The first effort to summarize all currently available research on the topic of assessment accommodations was published in May 1999 by the Mid-South Regional Resource Center (MSRRC) at the University of Kentucky. It is entitled, *A Summary of Research on Test Changes: An Empirical Basis for Defining Accommodations*, and written by Gerald Tindal and Lynn Fuchs.[1]

The research synthesis by Tindal and Fuchs was designed to provide empirically-based data to inform states regarding decision making for assessment policies. The authors recognize that assessment accommodations is a new and evolving area of knowledge.

This synthesis brief provides a succinct summary of the pertinent findings presented by Tindal and Fuchs.

Accountability Systems

Tindal and Fuchs ground their analysis of the research in a review of educational accountability, attributing the evolution of special education as a separate system to the use of the individualized education program (IEP). This has resulted in special education developing its own methods to account for outcomes. However, the authors note that serious technical problems exist in aggregating information across IEPs as an accountability measure. In addition, using such an accountability measure would have professional development implications that cannot be underestimated.

Almost all states have invested heavily in developing accountability systems that include statewide assessments. It is apparent that the widespread use of the IEP or any other separate indicator as an accountability mechanism for students with disabilities will not be widely accepted nor implemented in the foreseeable future beyond, perhaps, the needs of special education program evaluation.

It has been well documented, especially in the work of the National Center on Educational Outcomes (NCEO), that the inclusion of students with disabilities in large scale assessments has been minimal and extremely variable across states and districts. Participation has been based on ambiguous decision-making guidelines that are often influenced by perceptions of negative impact on system scores or emotionally-based impressions of negative effects of the test experience on the students. Even where such students were tested, their results were often excluded from accountability measures and/or from public reporting.

[1] A copy of the Tindal and Fuchs document can be obtained on the MSRRC web site: www.ihdi.uky.edu/msrrc

In spite of what has happened in practice, the current educational reform movement emphasizes accountability for the outcomes of *all* students. In addition, the inclusion of students with disabilities in general accountability systems is now supported in law. However, states and districts are encountering numerous dilemmas. Not the least of which is how to accommodate students with disabilities on assessments so that their knowledge and skills are measured without interference from the limitations their disability might impose on how they can express or demonstrate their achievement.

Terminology/Definitions

The 1997 Amendments to IDEA require states to include students with disabilities in their assessments and accountability systems with appropriate *accommodations* and *modifications*.[2] Agreement on the meaning of these terms has not yet been reached. In fact, one of the purposes of the Tindal and Fuchs synthesis was exploration of empirical support for the most commonly cited distinction between these terms. They conclude that a test change is an *accommodation* if it meets the following criteria:

- ❖ it does not alter the construct of what is being tested,
- ❖ it is based on individual need, and,
- ❖ it is effective for students who need the change and is not effective for others.

A test change becomes a *modification* when it does not meet these criteria. Briefly stated, a modification changes the nature of the construct being measured. It also is often stated that accommodations "level the playing field" by removing "construct-irrelevant" barriers to performance while preserving the meaningfulness of the scores for students with disabilities. In other words, students with disabilities should perform better with an accommodation and the accommodation should not benefit those without disabilities; a concept referred to as *differential effectiveness*.

Research Studies Included in Synthesis

A total of 114 studies were reviewed by Tindal and Fuchs. Each met the following criteria:

- ✓ The study examined a large-scale testing effort that reflected a broad measure of achievement (i.e., not a criterion test used as part of classroom instruction).
- ✓ The study documented a change in testing, either group or individual.
- ✓ The primary focus of the study was students with disabilities; however, some studies on students without disabilities were included if the outcomes had a bearing on test changes for those with disabilities.

The studies were reviewed using the National Center on Educational Outcomes's (NCEO) accommodation taxonomy.(See Appendix A for chart.) Most studies focused on only one test change, but some examined multiple changes. For each type of test change, the reviewed research is summarized and the quality of that research is analyzed. Tindal and Fuchs include an annotated reference for each study (type of adaptation, subjects, test used, and findings). The final section of the synthesis raises critical questions to address in future research on test changes.

This brief report summarizes the studies reviewed in each section of the synthesis, and highlights major findings for each category of change.

[2]IDEA Statute Sec. 612(a)(17); IDEA Regulations Sec. 300.138(a).

Summary of Available Research

Timing

Much of the research on timing covers college-age students with learning disabilities; however, some studies have begun to appear on the K-12 population with varying disabilities. Twenty-five studies were reviewed on this topic.

> *Research confirms that extended time may be an appropriate accommodation. Timed test conditions may not allow students with disabilities to reflect their full abilities, and may actually introduce error variance attributable to disability-related reasons.*

Setting

A change of test setting usually refers to the physical location where a test is administered. One study, however, addressed the variable of background music that altered the setting. While only three studies were reviewed on this topic, all were judged to be sound research and very credible.

> *Positive effects for students with disabilities were noted as a result of test setting change, while no such improvements resulted for students without disabilities.*

Computer Presentation

A distinction is made between computer-based testing (CBT) and computer-assisted testing (CAT) in the 16 studies reviewed in this area. CBT is the use of a computer to administer a conventional test. CAT is when the test content changes based on correct or incorrect responses as the student proceeds. There is almost no research on CBT with students with physical disabilities, even though this may be the only viable accommodation that provides access without modifying the meaning of the test. The CAT type of test was not reviewed in the Tindal-Fuchs synthesis. There is a need for research on the use of CAT because it is a tool that allows students to establish their own optimal item environment, and this could be an important accommodation device.

> *Research findings on computer-based testing must be interpreted with caution due to rapidly changing technology and the lack of large-scale studies.*

Familiar Examiner Presentation

The 10 studies reviewed in this area suggest that examiner familiarity may be a critical factor in the implementation of other accommodations that might require the use of persons unfamiliar to the student (e.g., extra time or smaller groupings). The implementation of these accommodations may inadvertently cause a negative effect on the student's performance.

> *Research findings favor the use of familiar examiners, typically the student's classroom teacher, to provide accommodations and administer the test.*

Multiple Focus Presentation/Response

The review of 31 studies by Tindal and Fuchs revealed that dividing test changes into either "presentation" or "response" was not always clear cut. For example, dictating to a scribe involves changes in both areas. Research in this area has the longest history; however, it is very complex to interpret because test changes have been done in packages. In addition, the positive effects on student performance have not always been differentially effective. Some of the research has been controversial. For example, Koretz's [3] analysis concludes that the use of some accommodations made the test biased; Trimble's[4] findings from the same data disputes this.

The use of large print or Braille and the reading aloud of math problems have consistently been shown to be differentially effective.

The Educational Testing Service (ETS) conducted research on accommodations used for college entrance exams (e.g., Braille, cassette recording, alternate recording systems, separate test locations, and extra time). Tindal and Fuchs note, however, that this research has limited applicability to students with disabilities since it focuses on college-bound individuals, a very small percentage of which have disabilities. The general findings from the ETS studies include the following:

Standard and non-standard administration of the Scholastic Aptitude Test (SAT) have comparable reliability.

Braille versions of the math test are <u>not</u> similar.

Scores from non-standard administrations are not systematically inflated or deflated and should be reported in the same manner as other scores.

Admission decisions need to be made with minimal effect from flagging of results from non-standard administration.

Studies on the use of test changes for the National Assessment of Educational Progress (NAEP) is a new addition to the research literature because accommodations have only recently been permitted. Issues around the maintenance of trend lines for future comparisons have been significant for NAEP.

Early findings from NAEP indicate that accommodating students with disabilities is likely to increase their participation, although the test may be different for this group than those in the general population due to the use of accommodations.

[3] Koretz, D. (1997). *The assessment of students with disabilities in Kentucky.* (CSE Technical Report No. 431. Los Angeles, CA: CRESST.

[4] Trimble, S. (1998). *Performance trends and use of accommodations on a statewide assessment.* Maryland/Kentucky State Assessment Series Rep. No. 3). Minneapolis, MN: NCEO.

Dictation to a Proctor or Scribe

Six studies were reviewed that examined dictation to a proctor or scribe during testing. However, little information was presented on the selection or training of the scribes or the rules used to change the speech responses into text that may have included a wide range of variance. The authors also note that, as voice recognition software becomes more sophisticated, policy makers may have to struggle with the extent to which such technology may represent an appropriate accommodation or provide an unfair advantage.

> *Dictation to a proctor or scribe appears to be uniformly effective in improving performance for students with and without disabilities.*

Alternative Responses

Six recent studies were reviewed in this area. The distrust of multiple choice tests and research on curriculum-based measurement has spawned a plethora of new formats and systems such as extended response items and portfolios. The authors comment that the IDEA mandate for alternate assessments will likely generate more research and development in this area.

> *Research done on alternative responses reflects the difficulties inherent in constructing comparable tasks, and there are many issues about research methodology that are still uncertain.*

Marking Responses in Test Booklet

Allowing students to mark responses in the test booklet is intended to eliminate the possibility that student responses on a separate "bubble sheet" are incorrectly aligned with the item presented in the test booklet. Four studies were reviewed that addressed this test change.

> *Marking in test booklets is an accommodation that shows no positive or negative results and it does not differentially benefit students with disabilities. At best, this accommodation is effective only for some individual students.*

Working Collaboratively with Other Students

This type of change is a natural extension for students who receive instruction in cooperative learning settings. Although the three studies reviewed on this topic question whether test outcomes are reflective of the individual or the group, this area of research contains only small studies done with few students who participated in groups in unknown ways.

> *The lack of vigor in experimental design in the studies leads to the conclusion that using collaborative groups to assess individual performance is likely to be unacceptable as an accommodation at this time.*

Using Word Processors

Tindal and Fuchs note that the ten studies reviewed from the literature on use of word processors is quite recent and needs to be replicated. Also, research to date has identified an intriguing number of variables that need to be studied. The only way to establish an empirical basis for decisions is to develop a systematic program of research. In using the computer in writing tests, all dimensions of the process need to be considered — the task, the process, and the judgment. It is critical to keep in mind that the goal of accommodations is to compensate for limitations brought on by the manner in which the test is administered or taken and not just to improve performance.

Findings on the use of word processors in writing tests are contradictory. Many variables need to be considered such as type of computer, use of features such as spell check, and student experience with the computer.

Using Calculators

The four studies reviewed in this area focus on the types of math problems because the effects of using calculators during testing is a function of problem type. It is generally accepted that calculator use for math problems that are designed to assess rote calculation is not appropriate and would invalidate any judgments of student proficiency.

Research indicates that using calculators during testing is not consistently helpful nor differentially effective.

Reinforcement

It is widely accepted that reinforcement is effective in maintaining behaviors; however, it is rarely used in testing situations. Twelve studies were reviewed in this area. Research on reinforcement during testing has been well conducted with classical group designs and a strong focus on students with disabilities.

Performance on standardized tests, whether aptitude or achievement, is greatly enhanced by reinforcement during the testing situation.

Instruction on Test-Taking Strategies

The four studies reviewed related to instruction on test-taking strategies are not recent and did not determine differential effects. Better research design and clearer definitions of treatments that isolate specific skills are needed.

Training in test-taking may help those with who lack experience in testing, such as students with disabilities. However, research has not focused on this population.

Instructional Level Testing

Instructional level testing, also referred to as out-of-level testing, usually involves the administration of items at a grade level earlier than the student's current placement. There is growing interest in this type of accommodation because of the emerging dilemma of how to assess students for whom both the general test and the alternate assessment are inappropriate. Only one study has been conducted and definitive conclusions are not appropriate.

However, given that computer assisted testing (CAT) focuses on item presentation at the student's instructional level, more research related to the use of computers as an accommodation will likely be done as CAT increases.

Observations

- It is unquestionably easier to demonstrate that a test change is an accommodation for a person with a physical disability than to prove the lack of connection between an accommodation for a person with a cognitive disability and the construct being tested.

- Although Tindal and Fuchs delineated the criteria to distinguish between a *test accommodation* and a *test modification*, perceptions continue to differ. For example, Braille is usually considered to be an accommodation, but some test companies maintain that the changes needed to make a test accessible through Braille often are significant enough to be considered construct changes.

- The importance of validating findings for students both with and without disabilities is vital in order to demonstrate that an accommodation does what it is supposed to do — eliminates an unfair disadvantage for students with disabilities. Tindal and Fuchs note that "validity itself has evolved from a purely measurement issue to a political and social issue" (p. 82).

Recommendations

- Teacher knowledge and perception of test changes must be improved in light of identified deficits that increase the error component of the outcomes.

- The process for creating policy should be anchored to an experimental rather than descriptive or comparative approach. Many more test changes are being recommended than we have data to support. Researchers and practitioners need to collaborate more effectively, and the research on test changes needs to be executed more carefully, on more diverse student populations, using different tests and with different decisions.

Next Steps

Plans are being made to load the data used for the Tindal-Fuchs synthesis onto the web site of the Mid-south Regional Resource Center so that it could be maintained and expanded as future research on the topic of test changes becomes available .

This report was supported in whole or in part by the U.S. Department of Education (Cooperative Agreement No. H159K70002). However, the opinions expressed herein do not necessarily reflect the position or policy of the U.S. Department of Education, and no official endorsement by the Department should be inferred.

U.S. Office of Special Education Programs

*Appendix A: Test Change Chart **

Setting

- **Administer the test individually in a separate location**
- **Administer the test to a small group in a separate location**
- **In a small group, study carrel**
- Provide adaptive or special furniture
- Administer test in locations with minimal distractions

Timing/Scheduling

- Use flexible - schedule
- Allow frequent breaks during testing
- **Extend the time allotted to complete the test**
- Administer the test in several sessions, specify duration
- Provide special lighting
- Time of day
- Administer test over several days
- **Provide special acoustics**

Response

Test Format:

- Increase spacing
- Wider lines and/or wider margins
- Graph paper
- Paper in alternative format (word processed, Braille, etc.)
- **Allow student to mark responses in booklet instead of answer sheet**

Assistive Devices/Supports:

- **Word processor**
- Student tapes response for later verbatim
- **Calculator, arithmetic tables**
- **Spelling dictionary or spell check**
- Typewriter
- Communication device
- **Alternative response such as oral, sign, typed, pointing**
- Brailler
- Large diameter, special grip pencil
- Copy assistance between drafts
- Slantboard or wedge
- **Tape recorder**
- Abacus

Presentation

- **Braille edition or large-type edition**
- Prompts available on tape
- Increase spacing between items or reduce items/page-line
- **Increase size of answer bubbles**
- Multi-choice answers follow questions down bubbles to right
- Omit questions which cannot be revised, prorate credit
- **Teacher helps student understand prompt**
- Student can ask for clarification
- **Computer reads paper to student**
- Highlight key words or phrases in directions transcription

Test Directions:

- **Dictation to a proctor/scribe**
- Signing directions to students
- **Read directions to student**
- Reread directions for each page of questions
- **Simplify language in directions or problems**
- Highlight verbs in instructions by underlining
- Clarify directions
- **Provide cues on answer from**
- Provide additional examples

Assistive Devices/Supports:

- Visual magnification devices
- Templates to reduce visible print
- Auditory amplification device, hearing aid or noise buffers
- **Audiotaped administration of sections**
- Secure papers to work area with tape/magnets
- **Questions read aloud to student**
- Masks or markers to maintain place
- Questions signed to pupil
- Dark/heavy/raised lines or pencil grips
- **Assistive devices-speech synthesis**
- **Amanuensis (scribe)**

***Note: Only bolded items were covered in studies of the Tindal-Fuchs synthesis.**

Source: National Center on Educational Outcomes (NCEO)

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

NOTICE

REPRODUCTION BASIS

☐ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☑ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").