

TABE Guidelines to Inclusive Testing Accommodations



GUIDELINES FOR INCLUSIVE TEST ADMINISTRATION

Introduction

Data Recognition Corporation understands the need to include all students in large-scale testing programs as part of the educational process. Federal, state, and local regulations or policies often stipulate that students be provided with appropriate accommodations during testing. The characteristics of the tested population are expected to change in response to the movement toward inclusiveness. Test publishers and those who desire to make valid interpretations of test results must consider how these changes affect the concept and practice of standardized assessment.

DRC supports the use of appropriate testing accommodations for examinees taking standardized tests and supports the efforts of decision makers to make valid and useful interpretations of test data.

In this section, DRC provides guidelines on the use and appropriate interpretation of the results of inclusive test administrations. These guidelines are intended to facilitate the valid interpretation of individual examinee results and valid comparisons of year-to-year and group-to-group summary data. They do not address the entire range of issues that must be considered when testing examinees who have Limited English Proficiency (LEP) or who are classified as English Language Learners (ELL). Agencies must become familiar with the specific requirements that apply to testing these examinees.

First, we present a general framework for reconciling standardization and accommodation in support of inclusive testing practices. Second, we discuss how to interpret criterion-referenced and norm-referenced test scores resulting from inclusive test administrations and present recommendations on the use of individual and summary results. Third, we detail a simple framework for classifying accommodations in terms of the potential effect on the appropriate interpretation of examinee test scores. This framework is illustrated using an arrangement of accommodations provided by the National Center on Educational Outcomes (NCEO) (*Thurlow, House, Boys, Scott, & Ysseldyke, 2000*). Finally, we list recommendations for constructive interpretations of aggregate results with regard to disabilities and accommodations. This approach provides relevant information for the use of test results, maintains examinee privacy, and avoids the unintended, negative consequences of labeling or "flagging" accommodated examinees.

1. A General Framework for Reconciling Standardization and Accommodation in Support of Inclusive Testing Practice

Standardization is a fundamentally important characteristic of educational assessments that are designed to support comparisons among participants. Historically, the definition of standardization in educational assessment has focused on compliance with uniform administration conditions. Requiring uniform administration conditions has resulted in the systematic exclusion of examinees for whom those conditions are not appropriate. The purposes of modern educational assessment now extend beyond examinee-to-examinee comparisons under uniform conditions, and the inclusion of all examinees in educational assessment has become highly valued and required by law. Furthermore, the interpretations to be made from assessment results have broadened, and the validity of these interpretations may be compromised by a requirement of uniform conditions. A reconceptualization of the principle of standardization is required to support the valid interpretation of results from inclusive test administrations.

DRC advocates and has adopted an approach to standardization that recognizes inclusiveness and accommodation as equally important, non-conflicting characteristics of modern assessment practice. The approach consists of four main principles. First, publishers of standardized tests should clearly define default conditions under which tests are to be administered, and these conditions should allow for broad participation by the vast majority of examinees. This principle discourages default conditions that are needlessly restrictive (e.g., short time limits, very small print), but also recognizes the importance of defining default conditions rather than leaving them to the discretion of administrators.

Second, changes to a publisher's default conditions will be necessary for some examinees to meaningfully participate in the assessment. Decisions regarding the use of such accommodations should be made by appropriately trained individuals familiar with examinee needs. These decisions should be documented in writing, such as in individualized education programs (IEPs), 504 Plans, or accommodation plans specific to LEP/ELL examinees. Under this principle, it is generally not appropriate for test publishers or policy makers to decree some accommodations as universally valid or invalid. Policy makers may exercise judgment regarding the treatment of scores arising from various accommodations.

The third principle is to define as standard the inclusive administration of assessments to all examinees who may meaningfully participate under either default or accommodation-specified conditions. As a result, examinees with documented needs for accommodations will participate in assessments under the accommodated conditions they experience in daily instruction. This inclusive definition of standardization guides the creation, publication, and norming of standardized tests at DRC, including TABE®. We believe that it supports a more meaningful set of interpretations from the results.

Finally, interpretation of the results of inclusive administration requires careful consideration of the targeted skills being measured, the nature and frequency of accommodations used, and the likely impact of the accommodations on reported performance. We elaborate on this principle in the next section.

2. Appropriate Interpretations of Test Results

In this section, we discuss appropriate interpretations of test results when there are examinees who take the test under accommodated conditions. That is, we focus on the validity of inferences to be made from the results of a test. When test administration conditions vary from the default conditions specified by the test publisher, the interpretation of test scores, both criterion- and norm- referenced, should take into account the actual administration conditions.

Criterion-referenced interpretations of test scores may be supported. They represent a fixed level of achievement that can be interpreted in terms of what examinees know and are able to do at a given score, either raw score, scale score, or performance level defined by a range of scale scores. For example, an examinee who achieves a performance level designated as "Proficient" on the Level L Pre-Reading test may demonstrate the following knowledge, skills, and abilities:

Proficient examinees demonstrate knowledge of symbol relationships in letters and words; demonstrate knowledge of sound/symbol relationships in letters; recognize initial or final consonant sounds or digraphs; and recognize short, long, and variant vowel sounds.

When an examinee achieves the "Proficient" performance level with the accommodation "extra time," for example, the testing conditions should be considered along with the knowledge and skills ascribed to the examinee. The interpretation may be that, given the particular raw score, scale score, or performance level, the examinee can demonstrate the knowledge, skills, and abilities cited above, with the accommodation "extra time."

Norm-referenced interpretations of test scores may also be supported. National percentile rank (NP), normal curve equivalent (NCE), and grade equivalent (GE) scores are examples of norm-referenced scores. These scores are interpreted in terms of an examinee's performance compared with the performance of a specified norm group (traditional or inclusive). When an examinee achieves a given norm-referenced score, say the 50th NP, on a mathematics test with the accommodation "examiner reads the test directions, stimulus material, and questions," the testing conditions should be considered along with the NP score. In this case, the valid interpretation is that the examinee who took the mathematics test, which was read aloud, performed as well as or better than 50% of the examinees in the norm group. In accordance with the principles set forth, DRC's inclusive norm group includes those examinees able to participate in the test administration with or without accommodations.

Given these interpretive guidelines, DRC recommends the following approaches for interpreting the results of inclusive test administrations.

Appropriate information for the interpretation of individual performance

An examinee who takes a test using accommodations should receive the same scale score referenced to the same norms tables as examinees with the same test performance achieved under default conditions. However, individual examinee results obtained using testing accommodations should be interpreted in light of the accommodation(s) used. As a supplement to such interpretations, classification summaries for three categories of accommodations are included in this section. These summaries are comprised of common accommodations categorized according to the effect on the appropriate interpretation of individual examinee results, as detailed in the following section.

Appropriate information for the interpretation of group performance

DRC recommends that summaries of results that are used for accountability purposes be presented in both aggregated and disaggregated forms. Aggregated results are summaries of results that include all examinees tested. These should be presented with the number and percent of examinees tested using accommodations so that the aggregated results can be interpreted with respect to changes in the use of accommodations across groups and years. Identifying the number and percent of examinees using accommodations provides valuable information. Disaggregated result summaries include only examinees who meet a specified criterion, such as examinees tested under the conditions defined as default by the test publisher. DRC recommends disaggregating results for examinees who take tests under default conditions and presenting the results separately from those for examinees who take tests under accommodated conditions. However, reports should never be presented for any group for which the number is so small that the confidentiality of examinee information would be violated (*FERPA, 1974*). Please note it is also important not to base inferences or important decisions on small numbers of examinees.

3. A Framework for Classifying Accommodations

On the following pages, DRC outlines a simple framework for classifying three categories of accommodations for use in reporting. To make the framework concrete, we classify into our categories each accommodation listed as common by NCEO. We place each assessment accommodation into one of three categories, depending on the potential influence on the appropriate interpretation of examinee test scores. The local agency may also use the classification criteria listed below to categorize any accommodations that are not shown here.

The following list of accommodations is not intended to be exhaustive, nor are the classifications of accommodations meant to be definitive. The classification of an accommodation may change for a particular test or subject area and may vary for different item types, such as selected-response and constructed-response items. The results of the limited research on the impact accommodations have on examinee performance are inconclusive. That is, the overall impact of accommodations on examinee scores may vary. Further, some accommodations appear to "boost" results both for examinees with and for examinees without accommodation needs. In addition, not all practitioners will agree with all of the classifications described. The purpose is to provide a framework for the classification of accommodations that may be used as policy and/or thoughtful application dictate. Both criterion- and norm-referenced scores are meaningful and useful when interpreted according to the following guidelines.

Category 1

Category 1 accommodations are not expected to influence examinee performance in a way that alters the interpretation of either criterion- or norm-referenced test scores. Individual examinee scores obtained using Category 1 accommodations should be interpreted in the same way as the scores of other examinees who take the test under default conditions. These examinee scores should be included in summaries of results without notation of accommodation(s).

Example of a Category 1 Accommodation: Take the test alone or in a study carrel.

This and similar accommodations comprise Category 1. These accommodations are not expected to affect the interpretation of individual examinee criterion- or norm-referenced scores. Both individual results and summaries can be treated in the same manner as those for examinees who do not use accommodations. Individual examinee scores can be interpreted directly, and examinee scores can be included in the aggregate score reports without special notation of the accommodations.

Category 1 Accommodations

Presentation

- Use visual magnifying equipment
- Use a Large Print edition of the test
- Use audio amplification equipment
- Use markers to maintain place

Response

- Mark responses in test book
- Mark responses on Large Print answer document
- For selected-response items, indicate responses to a scribe
- Record responses on audiotape (except for constructed-response writing tests)
- For selected-response items, use sign language to indicate responses
- Use a computer, typewriter, Braille writer, or other machine (e.g., communication board) to respond
- Use a template to maintain place for responding
- Indicate responses with other communication devices (e.g., speech synthesizer)

Setting

- Take the test alone or in a study carrel
- Take the test with a small group or different class
- Take the test at home or in a care facility (e.g., hospital) with supervision
- Use adaptive furniture
- Use special lighting and/or acoustics

Timing/Scheduling

- Take more breaks (Note: breaks should not result in extra time for testing or opportunity to study information in a test already begun)
- Have flexible scheduling (e.g., time of day, days between sessions), which should not result in extra time for testing or opportunity to study information in a test already begun

Category 2

Category 2 accommodations may have an effect on examinee performance that should be considered when interpreting individual examinee criterion- and norm-referenced test scores. In the absence of research demonstrating otherwise, test scores and any consequences or decisions associated with them should be interpreted in light of the accommodation(s) used.

Example of a Category 2 Accommodation: Use extra testing time for any timed test.

The accommodation "Use extra testing time" may have an effect on examinee performance, which should be considered when interpreting examinee scores. Category 2 includes accommodations that may affect the interpretation of individual examinee criterion- and norm-referenced scores. DRC recommends that individual examinee scores obtained with Category 2 accommodations be interpreted in light of the accommodation(s) used.

Category 2 Accommodations

Presentation

- Have directions read aloud
- Use a tape recording of directions
- Have directions presented through sign language
- Use directions that have been marked with highlighting
- Have stimulus material, questions, and/or answer choices read aloud, except for a reading comprehension test

- Use a tape recorder for stimulus material, questions, and/or answer choices, except for a reading comprehension test
- Have stimulus material, questions, and/or answer choices presented through sign language, except for a reading comprehension test
- Use communication devices (e.g., text-talk converter), except for a reading comprehension test
- Have computer presentation of a test that is not otherwise available for computer presentation
- Use a calculator or arithmetic tables, except for a mathematics computation test

Response

- Use graph paper to align work
- Use a spell checker, except with a test for which spelling will be scored
- For constructed-response items, indicate responses to a scribe, except for a writing test

Timing/Scheduling

- Use extra time for any timed test
- Take more breaks (Note: breaks may result in extra time for any timed test)
- Extend the timed section of a test over more than one day, even if extra time does not result
- Have flexible scheduling that results in extra time

Category 3

Category 3 accommodations are likely to change what is being measured and have an effect that alters the interpretation of individual criterion- and norm- referenced scores. This occurs when the accommodation is closely related to the knowledge, skill, or ability being measured (e.g., having a reading comprehension test read aloud). In the absence of research demonstrating otherwise, criterion- and norm-referenced test scores and any consequences or decisions associated with them should be interpreted not only in light of the accommodation(s) used, but also in light of how the accommodation(s) may alter what is measured.

Example of a Category 3 Accommodation: Use a calculator for a mathematics computation test.

The accommodation "Use a calculator for a mathematics computation test" may have an effect on examinee criterion- and norm-referenced scores and may change what is being measured. Consider the skills the examinee must demonstrate to solve the following problem with and without a calculator.

Subtract: 3,438 - 2,954 =

An examinee taking the item under default conditions, without a calculator, must demonstrate the ability to subtract and apply the concept of "regrouping." An examinee using a calculator does not need to apply the concept of "regrouping," but instead must demonstrate the ability to perform subtraction on a calculator. The skill measured by this item is clearly different for examinees using a calculator than for those who do not, and both criterion- and norm-referenced scores must be interpreted accordingly.

Most Category 3 accommodations are specific to the test content. DRC recommends caution when interpreting individual examinee scores obtained using Category 3 accommodations. Score interpretations should consider the accommodation-assessment combination and whether the accommodation(s) will change what is being measured.

Category 3 Accommodations

Presentation

- Use Braille or other tactile form of print
- On a reading comprehension test, have stimulus material, questions, and/or answer choices presented through sign language
- On a reading comprehension test, use a text-talk converter, where the reader is required to construct meaning and decode words from text
- On a reading comprehension test, use a tape recording of stimulus material, questions, and/or answer choices
- Have directions, stimulus material, questions, and/or answer choices paraphrased
- For a mathematics computation test, use a calculator or arithmetic tables that change the construct being measured
- Use a dictionary when language conventions are assessed

Response

- For a constructed-response writing test, indicate responses to a scribe
- For a test for which writing will be scored, use a spell checker
- Use a dictionary to look up words on a writing test

TABE Tip: The Category 1, 2, and 3 accommodations are shown in a table on the following page for your convenience.

Accommodations Classifying Table

Category 1 Presentation	Category 2 Presentation	Category 3 Presentation
<ul style="list-style-type: none"> ● Use visual magnifying equipment ● Use a Large Print edition of the test ● Use audio amplification equipment ● Use markers to maintain place 	<ul style="list-style-type: none"> ● Have directions read aloud ● Use to tape recording of directions ● Have directions presented through sign language ● Use directions that have been marked with highlighting ● Have stimulus material, questions, and/or answer choices read aloud, except for a reading comprehension test ● Use to develop recorder for stimulus material, questions, and/or answer choices, except for, except for a reading comprehension test ● Use a tape recorder for stimulus materials questions, and/or answer choices, except for a reading comprehension test ● Have stimulus material, questions, and/or answer choices presented through sign language, except for a reading comprehension test ● Use communication devices (e.g., text-talk converter), except for a reading comprehension test ● Have computer presentation of a test that is not otherwise available for computer presentation ● Use a calculator or arithmetic tables, except for a mathematics computation test 	<ul style="list-style-type: none"> ● Use Braille or other tactile form of print ● On a reading comprehension test, have stimulus material, questions, and/or answer choices presented through sign language ● On a reading comprehension test, use to text- talk converter, where the reader is required to construct meaning and decode words from text ● Have directions, stimulus material, questions, and/or answer choices paraphrased ● For a mathematics computation test, use a calculator or arithmetic tables that change the construct being measured ● Use a dictionary when language conventions are assessed

Category 1 Response	Category 2 Response	Category 3 Response
<ul style="list-style-type: none"> ● Mark responses in test book ● Mark responses on Large Print answer document ● For selected-response items, indicate responses to a scribe ● Record responses on audiotape (except for constructed -response writing tests) ● For selected-response items, use sign language to indicate responses ● Use a computer, typewriter, Braille writer, or other machine (e.g., communication board) to respond ● Use a template to maintain place for responding ● Indicate responses with other communication devices (e.g., speech synthesizer) 	<ul style="list-style-type: none"> ● Use graph paper to align work ● Use of spell checker, except with a test for which spelling will be scored ● For constructed-response items, indicate responses to a scribe, except for a writing test 	<ul style="list-style-type: none"> ● For a constructed-response writing test, indicate responses to a scribe ● For a test for which writing will be scored, use a spell checker ● Use a dictionary to look up words on a writing test

Category 1 Setting	Category 2 Setting	Category 3 Setting
<ul style="list-style-type: none"> • Take the test alone or in a study carrel • Take the test with a small group or different class • Take the test at home or in a care facility (e.g., hospital) with supervision • Use adaptive furniture • Use special lighting and/or acoustics 		

Category 1 Timing/Scheduling	Category 2 Timing/Scheduling	Category 3 Timing/Scheduling
<ul style="list-style-type: none"> • Take more breaks (Note: breaks should not result in extra time for testing or opportunity to study information in a test already begun) • Have flexible scheduling (e.g., time of day, days between sessions), which should not result in extra time for testing or opportunity to study information in a test already begun 	<ul style="list-style-type: none"> • Use extra time for only timed test • Take more breaks (Note: breaks may result in extra time for any timed test) • Extend the timed section of a test over more than one day, even if extra time does not result • Have flexible scheduling that results in extra time 	

4. Additional Reporting Considerations for Inclusive Administration

Score reports for examinees who take tests under accommodated conditions sometimes indicate or "flag" the conditions of testing to encourage valid interpretation of the examinee test scores. However, this practice may have the unintended consequence of labeling, which can stigmatize the examinee. There may also be legal and regulatory restrictions on flagging the scores of examinees who use accommodations. DRC recommends that agencies fully understand any legal requirements applicable to all aspects of their specific testing programs, including the use of accommodations.

Agencies using both individual examinee results and summary results must have some awareness of the specific conditions of testing. At the individual examinee level, agencies must interpret the results appropriately, necessitating a specific awareness of the testing conditions. To facilitate appropriate interpretation of individual examinee results, accommodation decisions and use should be well documented. Because examiners are directly involved with test administration and the use of test results, they are apt to be the best providers of information about the testing conditions and accommodations used. Such documentation includes IEPs, 504 Plans, or LEP/ELL team recommendations for both instructional and testing accommodation and details about the actual use of accommodations, in both instruction and testing.

Knowledge of accommodated administration conditions cannot be assumed for administrators who will be using the summary results. Those using the summary results typically are not concerned with the testing conditions for individual examinees, but they do need to know the numbers and percent of examinees using accommodations at the school or state level, for example. This is necessary to provide valid comparisons from year-to-year or from group-to-group, as well as to discourage the use of accommodations when they are not legitimately needed.

Test publishers should provide the necessary summary information while maintaining anonymity at the individual examinee level. The TABE CompuScan® answer documents allow for special coding to indicate examinee disabilities and/or types of accommodations. Agencies may assign accommodation

information to one or more of these special code spaces, and DRC can disaggregate the necessary information on score reports and/or research data files. Test administration conditions may also appear on individual examinee reports, but will be interpretable only by those familiar with the code layout.

5. Summary

The evolution of inclusive testing practices has challenged traditional definitions and approaches to standardizing and norming educational tests. As schools and agencies continue to adopt inclusive test administration as standard practice, it is appropriate for publishers of educational assessments to adopt standardization and reporting practices that are more inclusive and informative. The approach pioneered by DRC is reflected in the guidelines presented in this section and in the reports associated with assessments standardized at DRC under inclusive norming practices.

These guidelines can help educators provide examinees with appropriate accommodations and enhance the validity of the interpretations of results for all participating examinees who take tests under default or accommodated conditions. DRC is engaged in ongoing research to address the consequences of accommodations. DRC researchers have conducted studies examining the comparative effects of testing accommodations on the performance of examinees with and without disabilities. We apply item response theory (IRT) methods to examine whether and how items function differently for examinees with differing levels of ability and proficiency. We pay careful attention to the potential effects of testing conditions on test score interpretations and adhere to the Standards for Educational and Psychological Testing (*American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999*).

Even with an aggressive program of research, there is still insufficient empirical evidence concerning the relationships between accommodations and test results. DRC urges using the guidelines set forth in this section and exercising professional care in developing policy and interpreting test results.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. American Educational Research Association: Washington, DC.

Clarizio, H. F., & Phillips, S. E. (1992). A comparison of severe discrepancy formulae: Implications for policy consultation. *Journal of Educational and Psychological Consultation*, 3, 55-68.

Family Education Rights and Privacy Act (FERPA) of 1974, 20 U.S.C. 1232 (g).

McDonnell, L., McLaughlin, M., & Morison, P. Eds. (1997). *Educating One and All: Students with Disabilities and Standards-Based Reform*. Washington, DC: National Academy Press.

Thurlow, M., House, A., Boys, C., Scott, D., & Ysseldyke, J. (2000). *State Participation and Accommodation Policies for Students with Disabilities: 1999 Update (Synthesis Report 33)*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Developed and published by Data Recognition Corporation, 13490 Bass Lake Road, Maple Grove, MN 55311. Copyright © 2017 by Data Recognition Corporation. All rights reserved. This publication or portions thereof may be reproduced and distributed as needed for educational purposes only. TABE is a registered trademark of Data Recognition Corporation. All other brands and product names found in this publication are the property of their respective owners and are not associated with the publisher of this publication.

Rev. 10.26.17