



# **How ETS Works to Improve Test Accessibility**



*Listening. Learning. Leading.®*

Copyright © 2010 by Educational Testing Service. All rights reserved.  
ETS, the ETS logo, LISTENING. LEARNING. LEADING., GRE, TOEFL and TOEIC  
are registered trademarks of Educational Testing Service (ETS) in the United States  
and other countries. CBAL and PRAXIS are trademarks of ETS. 15761

[www.ets.org](http://www.ets.org)



# How ETS Works to Improve Test Accessibility

## Contents

<b>Introduction .....</b>	<b>4</b>
<b>Section 1: ETS Actions to Improve Accessibility.....</b>	<b>6</b>
<b>Section 2: Design of Accessible Tests—Building in Accessibility During Initial Design .....</b>	<b>9</b>
<b>Section 3: Development of Accessible Test Questions .....</b>	<b>20</b>
<b>Section 4: Development of Accessible Nontest Materials.....</b>	<b>28</b>
<b>Appendix A: Overview of Test Adaptations .....</b>	<b>30</b>
<b>Appendix B: References on Accessibility.....</b>	<b>41</b>
<b>Appendix C: Glossary .....</b>	<b>44</b>

## Introduction

ETS is guided by its mission “to help advance quality and equity in education by providing fair and valid assessments, research, and related services. Our products and services measure knowledge and skills . . . for all people worldwide.” In addition, ETS takes very seriously federal legislation providing that assessments be made accessible to persons with disabilities or that alternate accessible arrangements be offered.<sup>1</sup>

For many years, ETS has offered alternate test formats (ATFs) to individuals with disabilities, thus extending access to a broader group of test takers. (For a description of the range of ATFs, see Appendix A: Overview of Test Adaptations.)

The purpose of this document is to describe the work done at ETS to enhance the accessibility of our assessments and related products for test takers with disabilities. It provides practical guidance about:

- how, given their constructs, assessments and related products can be made as accessible as possible to most test takers, including those with disabilities who do not need ATFs
- how assessments and products that may need to be adapted can be made more amenable to adaptation as ATFs

In the assessment context, “accessibility” means ensuring that the test taker can interact appropriately with the content, presentation, and response mode of the test. To the extent reasonably possible (that is, consistent with the construct that the assessment seeks to measure), the content and format of assessments should allow all test takers, including those with disabilities, to demonstrate their mastery of the knowledge, skills, and abilities (KSAs) being tested.

The practices described in this document are intended to be applicable to most assessments, but they are most relevant to assessments designed for a general population that includes individuals with disabilities. Assessments designed solely for individuals with disabilities or those based on modified or

---

<sup>1</sup> See the Americans with Disabilities Act (ADA) and its amendments, the Rehabilitation Act of 1973, the Individuals with Disabilities Act (IDEA), and the Assistive Technology Act (AT Act).

alternate achievement standards mandated by federal legislation may require additional considerations. In addition, this document does not focus on accessibility issues related to test takers who are English Language Learners.<sup>2</sup>

Enhancing the accessibility of a test is a complex process. Considerations that would increase accessibility for one group may cause problems for another. Additionally, any assessment program is a complex entity, with a range of decision makers, affected persons, and other interested parties. For those reasons, application of the procedures suggested in this document will likely differ somewhat between ETS-owned programs and those programs to which ETS contributes but that are owned by other organizations.

In any context, ETS test developers and researchers are prepared to play an important role in looking for features of assessments that may impede accessibility and asking how the assessment can be designed or revised to improve accessibility. The need for adaptations cannot be eliminated entirely; the goal is to minimize the adaptations needed to make questions suitable for test takers using alternate formats as well as for the general population, while still preserving assessment of the construct.

Note: Many of the terms in this document are defined in the glossary that begins on page 44. Check the glossary for definitions of unfamiliar terms or acronyms.

---

<sup>2</sup> For additional information on accessibility for English Language learners, see the *ETS Guidelines for the Assessment of English Language Learners*.

## Section 1: ETS Actions to Improve Accessibility

Assessment specialists at ETS strive to provide the best measure of test construct within the constraints of a testing program while protecting the rights of all test takers, including test takers with disabilities.

### Background

For ETS-owned assessments, ETS provides accommodations for people with disabilities to the extent reasonably possible given the construct of interest. We at ETS also understand that accommodations may often need to be provided for client-owned assessments. ETS designs tests that reduce the need for accommodations and make it easier to provide appropriate accommodations by reducing sources of construct-irrelevant score variance for people with disabilities. At the same time, ETS seeks to provide the best measurement for the most test takers.

Unfortunately, these goals are not always compatible. For example, a test for learners of a foreign language may use videos as stimuli for a spoken response. For most test takers, using videos rather than only audio<sup>3</sup> stimuli offers a more realistic simulation of speaking with another person in the target language. However, videos present a disadvantage for a test taker who has a visual impairment and who does not have the benefit that the speaker's facial expression and posture and the surrounding visual context provide. Test developers attempt to reach the most appropriate balance between the goals of increasing accessibility and of providing the best measurement of the construct for the most test takers.

### Important Distinctions

For test questions (including stimulus materials), presentation modes, and response modes, ETS distinguishes aspects that are **essential** for measuring the intended construct, those that are **helpful** for improving measurement, and those that are merely **incidental** and that offer no important advantage.

---

<sup>3</sup> In this document, "audio" includes all auditory forms of test delivery, whether employing a reader (with or without a script), a recording, or computer-generated synthesized speech.

## **Essential Aspects**

An **essential** aspect is one that is **required** to measure the intended construct. For example, measuring the ability to interpret a graph (as opposed to a more general requirement to measure the ability to analyze data) requires that a graph be presented as stimulus material. Although test takers with disabilities will access the graph in various ways, the lack of a graph as a stimulus would make it impossible to measure the intended construct. Such aspects must be retained for test validity even if they present accessibility barriers, and accommodations may be supplied as long as the essential aspect is retained. For example, a raised-line drawing of a graph may be used in place of a flat drawing if reading a graph is an essential feature of the construct and the test taker is blind.

## **Helpful Aspects**

A **helpful** aspect is one that improves the ability to measure the intended construct. It is not essential because the construct could still be measured without it. For example, using pictures as stimuli for speech in a test of English for nonnative speakers is a helpful technique for most test takers, but pictures are inaccessible for test takers who are blind. In deciding whether to retain helpful aspects that present accessibility barriers, test developers weigh the overall advantages of these aspects. Do they help make the assessment more discriminating, more efficient, more informative, more interesting, more realistic, more reliable, more thorough, less time-consuming, or result in more valid testing for most test takers? If so, do these advantages sufficiently outweigh the accessibility barriers for certain test takers?

## **Incidental Aspects**

An **incidental** aspect is one that could be removed or revised without significantly harming the ability to measure the intended construct and without lowering the quality of the test question. For example, even though the presentation of a graph may be essential, certain aspects of the graph that cause difficulty for people with particular disabilities (such as color coding of the lines) may be incidental.

## **Applying the Distinctions**

Sections 2 and 3 will expand on how test developers and test design teams apply the accessibility policy distinctions in designing new assessments and in carrying out the day-to-day work of developing test material. At the



design stage of test development, a precise definition of the construct with clear boundaries between included and excluded knowledge, skills, and abilities (KSAs) enables test designers to distinguish among essential, helpful, and incidental aspects. Defining the construct will guide test design teams in thinking about whether to retain a helpful aspect that may hinder accessibility.

- How important is the advantage provided by the helpful aspect?
- Is there some more accessible way to gain the same or almost the same advantage?
- Are there reasonable accommodations that allow measurement of the construct for people with disabilities?
- How big a problem does the helpful aspect present for people with disabilities? (The more problematic the helpful aspect, the less likely it is that it should be retained.)
- How much of the test is problematic for people with disabilities? (The smaller the portion of a test that is problematic, the more likely it is that the helpful aspect should be retained.)
- Is there a comparable, more accessible question type that can be used in the new form?
- For programs that adapt existing forms for alternate test format use, can problematic questions be deleted from an existing form while still allowing comparable measurement of the construct?

Beyond the design stage, test developers draw upon essential aspects to measure the test's construct. They also justify the use of helpful aspects and attempt to avoid incidental ones. The various question and test reviews at ETS, particularly fairness reviews, include an evaluation of the accessibility of questions and stimuli.

## Section 2: Design of Accessible Tests—Building in Accessibility During Initial Design

This section will address the following considerations:

- Accommodations—Of the accommodations typically approved for test takers with disabilities, how do design teams decide which are compatible with the test construct?<sup>4</sup>
- Question types—Are the question types under consideration accessible as is, and if not, are they adaptable?
- Presentation—How can the presentation of the assessment be enhanced for most test takers, including those with sensory or learning disabilities?
- Repurposing an existing test—What can be done to existing tests when they are repurposed?

### Rationale

When a new test is being developed, or when an existing test is being redesigned or repurposed, ETS strives to consider accessibility throughout the process, from the time the test construct is defined until the questions are written, reviewed, and assembled into tests. Doing so helps to ensure that projects are scoped and scheduled appropriately.

The initial stages of development of a new test or of evaluating a test for repurposing involve defining (or reconsidering) the construct. The term “construct” is used to refer to all of the knowledge, skills, abilities, and other attributes that a particular test is intended to measure (the KSAs). As the construct definition is evolving, members of the design team consider the construct in relation to test takers with disabilities and the accommodations commonly used to make tests accessible for them. Design teams that lack experience with accessibility concerns for individuals with disabilities seek advice from test development staff and other professionals with appropriate knowledge and experience with test and question adaptation.

---

<sup>4</sup> Decisions about an accommodation’s compatibility with the test construct apply regardless of what disability a particular test taker might have. (At ETS, decisions about how to accommodate individual test takers and particular disabilities are the responsibility of the ETS Office of Disability Policy and/or the client.)

Because the constructs of most ETS postsecondary tests involve some aspects of cognitive ability, certain cognitive disabilities may not trigger testing accommodations at the postsecondary level. These disabilities may be considered at the K–12 level because individuals with cognitive disabilities are part of the population being educated and so testing mandates apply. For disabilities such as intellectual disability, assessment may be addressed in a different manner (e.g., with small group or one-on-one administrations).

## **Determining Which Accommodations Are Compatible with the Test Construct**

In planning for accommodations, the design team focuses on the construct-appropriateness of a given accommodation for a particular test.<sup>5</sup> For example, “read aloud” or “audio” may be inappropriate for some reading tests because decoding text from symbols (e.g., determining that “c-a-t” means cat) is relevant to their construct. It is important to determine whether a particular skill, such as decoding, is not only *relevant* to the construct but is in fact *so important* to the construct that a relevant accommodation like read aloud or audio is inappropriate.

(See Appendix A for more information about accommodations that involve test adaptation.)

Defining the construct precisely will help the team determine whether a construct-relevant accommodation is allowable.<sup>6</sup> Making decisions about the accommodations can in turn help fine-tune the definition of the test construct; the process is iterative. The design team may decide to keep helpful but nonessential components in the test for the general population but remove them from certain test forms as required for accommodations. The team then determines the relative weights of various KSAs measured in the test. The final result is a well-articulated construct definition and a clear

---

<sup>5</sup> It is helpful for test developers and other design team members to recognize how people with relevant disabilities access or interact with material, so as to understand how best to adapt the material and how to evaluate the construct and possible accommodations in light of those modes of interaction. In this section, wherever the word “accommodation” is used, it is meant to include only those accommodations that are construct relevant. Such accommodations involve primarily, but not exclusively, adaptations to tests and questions.

<sup>6</sup> In K–12 programs, this process will involve discussions with client review committees. Sometimes construct definitions are a matter of state policy, and existing documentation supplies information about construct definitions and question specifications.

understanding of the relationship between the construct and each of the typical construct-relevant accommodations.

Here are two examples:

### **Example 1: Reading Test**

The construct for a particular reading test might include one or more of the following abilities:

- understanding written text in English (or in another language)
- understanding the type of language typically used in written texts
- decoding text from print
- decoding text from symbols (whether print or braille)

Before describing the construct as essentially involving print decoding, test-design teams consider the impact on test takers with print-related disabilities. If print decoding is essential (as in language tests using non-Roman alphabets or some occupational tests), individuals with visual impairments or decoding disabilities can be expected to achieve lower scores; braille or audio would be inappropriate. However, in many cases print decoding is a part of the construct but not an essential part. In such cases, formats other than print (e.g., braille) or formats that remove the need for decoding (e.g., audio) are construct compatible.

### **Example 2: Listening Test**

Listening tests are commonly included in language-proficiency assessments. The construct of a potential new listening test might include one or more of these abilities:

- understanding spoken text by listening to an audio recording of someone reading
- understanding the type of language typically used when speaking
- understanding language in real time, without the ability to review, such as in a classroom, lecture hall, or telephone conversation

The KSAs above are relevant to decisions about accommodations and adaptation issues such as:

- whether listening passages can be played more times than normally permitted
- whether a listening section can be omitted
- whether a written transcript can be provided in addition to or instead of a listening section

Assessments with audio content include music tests, language-proficiency tests, some professional-licensure tests, and reading tests that measure such processes as oral comprehension, following directions, or story recall. Listening components of music tests are often essential to measuring the construct and as such can be difficult or impossible to adapt for individuals with hearing losses. Music tests also present adaptation issues relevant to visual disabilities: musical scores are difficult to enlarge or have brailled and are often impossible to describe useably in a reader script.<sup>7</sup>

In language proficiency and reading tests, the situation is also complex. Hearing is an essential part of the construct in some listening tests. In others, the construct involves the type of language typically used in spoken language rather than the delivery mode (speech).<sup>8</sup> Separate yet somewhat parallel issues apply to the assessment of speaking.

If actually hearing the text is not essential to assessing the construct, test developers consider the test's purpose and whether it is compatible with the test construct to deliver material of this sort in written form.

Note that the goal is to be able to deliver accessible forms of the test, not necessarily to make every individual question adaptable for use with the accommodation. For programs that assemble test forms earmarked as alternate test forms, for example, it helps to have an adequate number of

---

<sup>7</sup> Adaptations of music tests for individuals who are hard of hearing are infrequent but extremely complex—for example, transposing a listening section an octave downward for someone with a high-frequency hearing loss. These difficulties can often be overcome, depending on the nature of the approved adaptation and on specific features of the test.

<sup>8</sup> The TOEFL® iBT provides an example of how the definition of the construct can vary from section to section of the same assessment. In the Listening comprehension section, the construct is defined as comprehending the target language by listening to spoken text; therefore, this section cannot be adapted by providing test takers with written transcripts of the stimuli. However, the skill tested by audio elements in the Speaking and Writing sections is defined differently: as comprehending the type of language typically used when speaking. Therefore, the audio portions of these sections can be adapted for test takers who are deaf or hard of hearing by providing transcripts of the audio stimuli.

accessible questions available for assembling and delivering forms that meet the test specifications and that are adaptable for the appropriate accommodations. For programs that do not assemble such forms but instead adapt existing forms on demand, it may be possible to remove nonadaptable questions without seriously impairing measurement of the construct.

## **Selecting Appropriate Accommodations**

Once the team has analyzed the KSAs that the new test is designed to measure, it then determines which accommodations can be made. These may include:

- ✓ Adjustments to presentation mode: braille,<sup>9</sup> audio (recorded, prepared for live reader, or computer-voiced), large print, magnification, tactile or enlarged figures, paper test (if the test is ordinarily computer based), written script of auditory component,<sup>10</sup> oral interpreter for auditory component
- ✓ Adjustments to response mode: computer with screen reader for test taker–written responses, voice recognition for test taker–written responses, scribe, assistance in operating recording equipment, large-print answer sheets, recording answers in test book, text-to-speech software, including text readers or screen readers<sup>11</sup>
- ✓ Omission of portions of test (e.g., omission of a speaking section for a test taker with a speech disability)
- ✓ Testing aids: calculator<sup>12</sup> (including large-display or talking calculators), abacus, spell-checker (typically, a program-approved model that does not include thesaurus or dictionary capabilities)
- ✓ Adjusted administration: extended time, ability to replay audio portions of tests, extra breaks, testing on multiple days, separate rooms, individual or small group administration

---

<sup>9</sup> Those developing reading tests for the early grades need to keep in mind that braille readers may be learning to read words in a different order because of the nature of braille.

<sup>10</sup> For some ETS foreign-language tests (and some others), providing a script does change the construct, since understanding spoken language auditorily is an essential element of the construct.

<sup>11</sup> For ETS-owned tests, use of a screen reader is permitted only on a limited basis because it may not be reasonably possible to create a fully screen reader–optimized test, especially if scientific or mathematical symbols, equations, and/or graphics are involved.

<sup>12</sup> A calculator and an abacus may be considered construct neutral when they merely help test takers make calculations more quickly. In addition, for some test takers (for example, those who are blind), an abacus may be considered the equivalent of scratch paper.

The most common testing accommodations do not involve any changes to questions or tests. There are, however, instances in which even these common testing accommodations can have an impact on the test itself. (For example, the design of audio aspects may be affected by the use of extended time if a standard response time is built into the recording.) In addition, some test constructs, such as reading fluency, are so time dependent that determining whether extended time is a permissible accommodation requires careful consideration. Does the construct of fluency require responses of some minimum speed? If so, providing more time is not appropriate. If fluency means clarity or accuracy of response rather than speed of response, then it is acceptable to provide more time for response.

## **Selecting Accessible and Adaptable Question Types**

Once the construct and test framework have been determined, test designers typically identify or develop appropriate question types to measure the construct. In considering question types for an assessment, the design team tries to keep in mind the accommodations that have been determined to be appropriate for the test. The team selects question types that are likely to be amenable to construct-appropriate adaptation, or it develops a plan for assembling alternate test forms that use only questions that can be adapted.

To the extent feasible and acceptable to the program, state, or other test owner, the design team solicits feedback from appropriate outside groups (e.g., content experts, users, other stakeholders) when a new assessment or question type is being developed and before it is first used for formal data gathering. The following represents an ideal list of ways to select question types appropriate for individuals with disabilities; often it is not reasonably possible to follow every step. When possible, design teams consider the following steps and choose among them as appropriate:

1. For each question type, develop accessible versions of an appropriate range of questions. Try to develop more than one way in which the questions could be made accessible. In doing so, it may be helpful to consult with experts in developing accessible questions.
2. Identify and work on closing potential holes in the pool of accessible questions so that whole forms of adaptable questions meeting the required statistical and content specifications can be assembled.



3. If allowed, share drafts of accessible versions of questions with outside experts (such as content experts who work with people with disabilities and experts with experience in producing or working with adaptations) to get feedback and suggestions for improving accessibility. It may be helpful if some of these experts are themselves people with relevant<sup>13</sup> disabilities.
4. Try out the accessible versions of the question types. Question tryouts may take several different forms, ranging from one-on-one interviews, through small-scale pilot tests, to large-scale field tests. The approach may be determined by cost and time considerations as well as the availability of qualified individuals familiar with the field being measured.
  - One-on-one interviews with test takers who have been administered the questions can provide much useful information. These interviews can take the form of informal debriefings after test takers have completed the tasks or more formal cognitive laboratory activities where test takers are interviewed either while they are answering the questions or afterward. Focus groups are another method of learning about the accessibility of sample questions, especially when interviews and cognitive laboratory activities are inadequate, for example with test takers who are nonverbal.
  - Small-scale pilot tests may also be used to obtain information on how test takers respond to the questions. Because these samples may not be fully representative of the test-taking population, the statistics provide only a gross measure of whether questions are accessible.
  - Programs with adequate resources may administer large-scale field tests to a representative sample of test takers that may include people who use each of the accommodations that have been developed. (Often even a small sample, for example of braille readers, can provide useful information.)
  - A test-taker survey or other feedback opportunity might be provided to participants in the pilot test. Individual comments can provide valuable insights into accessibility. See Laitusis (2007), included in the list of references in Appendix B, for more information.

---

<sup>13</sup> A relevant disability is one that decreases the likelihood that a test-taker's score will appropriately reflect his or her knowledge, skill, and ability; for example, blindness is a relevant disability for a test involving numerous photographic stimuli. However, for most tests, confinement to a wheelchair is not a relevant disability, since it primarily affects access to the test site but not access to the test itself.

- Analyze the results of the pilot or field test and make determinations as to what types of questions are appropriate to use and in what form. The pilot test results can suggest ways of making the questions generally more accessible for the target population and can help test developers to determine which formats to restrict in an accommodation.
5. If a given type of question is not amenable to accommodations, test designers consult with statistical analysis staff to determine if it is possible to generate a valid score when such questions are omitted or are replaced with other, more easily adaptable questions for alternate test formats.
  6. Plan to have a pool of substitute questions available in case any questions are found to be inaccessible.

Prudent decisions made at the test-design stage can save time and minimize costs later as well as enhance the accessibility of the final product.

## **Making the Presentation Accessible**

Most of what makes an assessment visually appealing and legible for special populations also works for the majority of the population at large. For example, everyone benefits if a graphic has a heading and there are labels for the key parts of the image. To the extent feasible, design teams strive to use layout, formatting, and typography that make the test clear. Well-designed headings and graphics can help the typical test taker recognize the relative importance of information and the order in which it should be considered. Numbered or bulleted lists for directions and other materials can be comprehended more easily than paragraphs of text.

Design teams may review available research and consider the following aspects of presentation:

- font type (e.g. sans serif)
- font size
- font style
- page justification (flush left is preferable)
- contrast
- line length

- leading
- conventions for graphics and associated text
- splitting of words between lines or of text between columns and pages
- format of headings
- screen resolution
- response format
- scrolling issues that involve set leaders, tables, or graphs
- split screens
- amount and location of blank space on the page or screen
- method of delivering instructions via computer (text and audio)
- timing and exit instructions on the computer
- for tests with audio components, the appropriateness of providing instructions simultaneously in text and audio
- for computer-delivered tests, the feasibility of leaving instruction and directions screens untimed and of allowing test takers to exit the screens manually when they are ready
- any other aspects of test delivery that can be made equally applicable to tests given with and without accommodations, including what the directions state about response modes and timing

Decisions about these features will be affected by the characteristics of the target population.

See Section 3 and Appendix A for more information about visual accessibility.

## **Repurposing an Existing Assessment**

An assessment is considered “repurposed” if it is to be used for purposes or populations other than those for which it was previously designed (e.g., using a test for placement that was originally developed for admission, or vice versa). Test purposes include, but are not limited to, accountability, licensure, admission, placement, scholarship awards, summative assessment of classroom instruction, diagnosis and feedback, measuring educational outcomes, and employment.

Repurposing requires reevaluating the test's construct and the construct's interaction with likely accommodations. Changing the purpose of a test can also change the degree to which accessibility for a given population is mandatory. For example, some tests with no consequences for individual test takers may have been useful for their original purpose, even if they were inaccessible to test takers who are blind.<sup>14</sup> However, once such a test is repurposed so that individuals have a stake in the outcome, blind test takers should be included to the extent that the construct permits.

Repurposing a test could require revising the accommodated forms that were originally developed. If the construct has changed, the appropriateness of certain accommodations may have changed as well; and if the population has changed, assumptions about the appropriateness of question types and individual questions may need to be examined. Thus, repurposing a test is very much like designing a new test: the construct, the ways of measuring it, and the characteristics of the target population should ordinarily be evaluated.

---

<sup>14</sup> However, if the test's purpose is to gather information about the achievement of all students in a certain grade or jurisdiction, even if the test has no consequences for individual students, the test does not serve its purpose if it is inaccessible to some students.

## Section 3: Development of Accessible Test Questions

Most day-to-day test development work is done on programs for which the general requirements—including practices related to accessibility—are well established. The guidelines in this section are meant to assist in enhancing the accessibility of test questions that are written and reviewed for such ongoing programs.

In working to enhance the accessibility of test questions, test developers consider two general goals:

- to develop questions that can be used by a wide range of test takers without the need for adaptation
- to develop questions that are amenable to adaptations when appropriate

This section will discuss ways to improve overall test validity by enhancing the accessibility of test questions. It is not meant to be prescriptive or to imply that questions and test content must be modified in a way that would compromise measurement of the construct or that would go against established policies.

More discussion of these issues can be found in articles listed in Appendix B.

### Learning About the Program

Assuming that test developers know how a particular assessment construct has been defined, an important next step in developing questions is to learn about the test features that are relevant to the accessibility of the assessment. For example, are graphics used? If so, what kinds? Are there any restrictions on the content, reading load, or vocabulary in the questions? Are alternate test formats or special test forms to be produced? If so, what kinds? What are their most important features? What issues would make the assessment less accessible or present challenges when it is adapted into alternate formats? (Information about typical adaptations and how they are created is provided in Appendix A.)

At ETS, programs with very large expected volumes may provide alternate test formats as part of their inventory. Such programs often have the advantage of being able to stipulate in advance which test forms will be

adapted for persons with disabilities. Other programs (often those with small volumes) provide accessible test formats on demand, which may require that assessment staff work with material that is difficult or impossible to adapt. In either case, it is not always possible to predict whether any given question will be used in an alternate format or on a test form taken by a person with a disability who takes a regular format test.<sup>15</sup> These facts underscore the desirability of working to enhance the accessibility of all test questions to the extent reasonably possible.

## **Making Informed Judgments**

ETS works to make tests valid and accessible for test takers with disabilities without compromising validity for the general test-taking population. Central distinctions are those between aspects of an assessment that are essential, helpful, and incidental. The most challenging decisions relate to those aspects of a test question that are classified as helpful, as such aspects call for a judgment that balances potentially conflicting interests.<sup>16</sup> Test developers who have inherited the job of developing questions for an existing test should consider eliminating incidental aspects that make questions less accessible. It may be useful for them to consult with individuals experienced in creating alternate test formats and to take into account client preferences as well as program guidelines and history.

Certain tests have auditory, visual, or physical components that are essential to their constructs and/or test specifications; for such tests, it would be inappropriate to change those components if doing so would compromise the construct. For example, many science assessments must use a high proportion of graphics to satisfy content specifications. The responsibility of the test developer in science is to ensure that the graphics are presented in a manner that balances the need to assess the construct with the need, to the extent reasonably possible, to be accessible.

---

<sup>15</sup> Not all test takers with disabilities request (or require) accommodations. A test taker with a hearing loss may have difficulty with questions that deal with sounds, phonics, etc., but may not request an alternate test format because he or she is not expecting to encounter any such questions. Most test takers with learning disabilities do not use alternate test formats but may have accessibility issues with some questions or question types.

<sup>16</sup> Accessibility considerations begin at the time assessment frameworks and specifications are being developed, if possible. When existing frameworks and specifications do not clearly address accommodation issues, additional attention on the part of the test developer is appropriate so that these issues are covered at the question-development stage.

For many skills or question types, a test question that is amenable to adaptations may measure the skill or knowledge as well as one that is not amenable. If the test developer knows in advance that a particular test form is destined to be used in alternate formats, he or she may select appropriate components when developing the form.<sup>17</sup>

## **Textual Content**

A resource that provides general guidance on making the language of an assessment accessible is “ETS Guidelines for Using Accessible Language in Tests,” which appears in the *ETS Guidelines for Fairness Review of Assessments* (available at [www.ets.org](http://www.ets.org) at no charge). The accessible-language guidelines were developed to minimize construct-irrelevant variance by helping test developers make the language of tests as accessible as possible within the bounds of the construct being assessed.

The following concerns relate to reading material that will be used with alternate test formats. The advice is meant to highlight ways to write and select questions that will facilitate the work of those who adapt test forms. (See Appendix A for more information.)

## **Stimulus Materials**

1. Stimulus passages may have complex features that are not easily scripted for a reader or for audio (unless such features are essential to test). Such features may include:
  - extensive dialogue
  - unusually long paragraphs or other complex structures
  - words that are difficult to pronounce
  - footnotes
  - a number of visual images or extensive descriptions
  - sidebars

---

<sup>17</sup> For example, extremely long passages, questions with multiple parts, dictionary pages, and multiple or complicated graphics can turn two facing pages into six to ten braille pages, making the task of relating the parts much more difficult for the braille user. Test specifications may not require that such materials be included.

If a particularly complex stimulus passage is not required by the construct, test developers may consider modifying the passage appropriately or substituting another passage.

2. Some stimulus passages depend on sensory knowledge and may disadvantage test takers with visual impairments or hearing loss. Judgment is important in this area: test takers who are blind, for example, are used to passing references to color or vision, so a key concern is whether experience with vision or hearing is necessary to understand a passage and to answer related questions. A passage or question about the life of artist Georgia O’Keeffe will likely be more accessible than one describing detailed aspects of her artwork.

Similarly, a test taker who is deaf is ordinarily more likely to be able to respond to questions about a reading passage on the life of Mozart but not to questions about certain technical aspects of his compositions. A question or passage that goes into detail about melody, harmony, or rhyme may be disproportionately problematic for test takers with hearing losses.

Even a question that does not depend strictly on sensory knowledge might assume detailed familiarity with such knowledge. An accumulation of such questions could overwhelm some test takers with disabilities.

### **Stems and Keys**

1. When it is practical to do so, questions are worded concisely to avoid long and complex stems, options, or descriptions. In addition to generally increasing reading load, these complexities are burdensome when repeated in any form of audio.
2. Where appropriate, clear signals are provided that help the test taker direct her or his attention (“in the first sentence” or “the graph shows,” for example). Note that some state programs do not allow any reference to graphics in the stem.

### **Directions**

1. All test takers benefit when directions are clear and straightforward.
2. Test takers also benefit when numbered or bulleted lists are used for directions and other materials that can be better comprehended in list



form (however, bullets may not be clearly conveyed in an audio adaptation).

## Visual Content

1. Some test takers (such as those who have low vision, a color vision deficiency, or cognitive impairments) may not easily perceive certain features that are included in figures in regular test forms; some figures may be problematic in the braille, audio, or large-print formats. Possibly problematic features may include:
  - complex visuals, or visuals that are not in a formulaic or standardized format (e.g., cartoons, photographs, or maps)
  - crowding of details<sup>18</sup>
  - labels (e.g., on maps) that are too small or that are irrelevant
  - multiple degrees of shading
  - text, including labels, printed on shaded areas
  - labels that are vertical, slanted, or anything other than horizontal
  - labels or other text in italics, handwriting, or decorative typeface
  - cluttered graphs (If consistent with the construct, test designers may use tables, grids, or text descriptions instead, make the graph less cluttered, or convert one cluttered graph into a few less-cluttered graphs.)
  - drawings or pictures of three-dimensional solids<sup>19</sup> (Some three-dimensional objects can be described successfully, but tactile-graphic representations are seldom decipherable or interpretable. Although it is rarely feasible to do so, actual three-dimensional models may be used.)
2. Visuals located in the middle of paragraphs may be inaccessible for certain test takers.

---

<sup>18</sup> In general and where reasonably possible, test developers try to avoid complex details or remove them from existing tests if they are not essential in the alternate test format. If complex details cannot be avoided, then a key or guide is provided if reasonably possible. Other strategies for mitigating the issue may raise questions related to valid measurement.

<sup>19</sup> Such pictures may be essential for some tests (e.g., three-dimensional representations of molecules).

3. Visuals in response options can also be problematic. (Although note that this practice is essential in some science, music, and mathematics tests.)
4. Illustrations that accompany reading-comprehension passages can be distracting. Ideally, illustrations are used only if necessary to test the construct. However, many K–12 programs require reading passages to have the same kinds of illustrations or decorative features that appear in books or magazines; the purpose is to enhance interest or to provide a realistic reading experience. Merely decorative elements can be omitted for visually impaired test takers. However, if such decorative elements also serve a minor but informative function such as “scene-setting,” a brief piece of text can be added (e.g., “The following story takes place in a classroom . . .”).
5. Pictures of student work (drawings, diagrams, etc.) that cannot usefully be described are often inaccessible to certain test takers. These pictures may be used when they are essential in measuring the construct (as in some teacher-licensure tests) but can often be omitted from tests administered to test takers who are blind.
6. Figures that can be described clearly and concisely are preferred.
7. For labels, a legible sans serif font, such as Helvetica, Verdana, or Arial, is recommended. Labels are usually arranged in a clockwise direction.
8. To the extent possible, the use of all capital letters for extensive text (for example, in directions) is avoided.

## **Audio Content**

1. Recorded stimuli are clear and reasonably paced for accessibility. (However, some test programs require authentic stimuli.)
2. Audio materials are prepared in a way that will facilitate the use of extended time, if extended time is compatible with the test construct. Test developers consider what part of the time gets extended (listening and/or response time) and whether scoring rubrics will have to be adjusted for constructed-response questions.<sup>20</sup>

---

<sup>20</sup> For tests with constructed-response questions, supplements to the scorer-training materials are often developed so that responses to the accommodated version of the test can be appropriately scored.

3. Test developers consider whether, given the construct, a written transcript and/or repetition of the stimulus are permissible. If so, content is supplied in a way that will facilitate making those adaptations when needed. For example, to the extent possible, scripts are retained and questions are presented on separate audio tracks so that an individual track can be repeated if permitted.

## Question Format

1. Test developers attempt to limit the use of complex question formats such as those with Roman numerals (in which answer choices include "I only," "I and II only," etc.). This question format is difficult to adapt for audio.
2. Open stems can be challenging for some test takers to deal with in audio format. However, if there is a logical syntactic break at the end of the stem, or if the stem and options are quite simple, an open stem may be just as good as or better than a closed stem. Compare, for example, these two stems:

Which of the following did Martin Luther King, Jr., tell people to do in order to achieve his goals?

To achieve his goals, Martin Luther King, Jr., told people to

3. Negative stems are used sparingly and only when required for valid measurement. Many testing programs use capital letters or other methods of emphasis<sup>21</sup> (e.g., "EXCEPT" or "NOT") to call attention to the negative stem. Some state programs prefer italics; others do not allow the use of any negative stems.
4. Extensive use of line and paragraph references, especially references to multiple portions of a stimulus within an item, can pose problems. Such references are helpful to sighted individuals but can lead to complications when adapting tests for audio, braille, or large print. In addition, these references can change in different versions. (Many programs, however, require the use of line numbers.)

---

<sup>21</sup> All-caps format is less legible than mixed case or lowercase type. Therefore, bold, underlined, or bold and underlined text may be preferable ways of conveying emphasis, especially on large-print tests.

5. Test developers try to follow the following guidelines regarding labels and symbols (unless content dictates otherwise):
- Using letters that are clearly distinguishable by sound (e.g., avoiding the use of *S* and *X* in the same question, unless it is a question that must use *S* and *X* for content reasons).
  - Using letters that are easily distinguishable from each other by sight (e.g., avoiding the use of *O* and *Q* in the same question).
  - Unless they are essential in testing the construct, avoiding invented or nonstandard symbols that have no equivalent in braille and symbols whose meaning or pronunciation is not obvious. These may be difficult to use in audio.

## Layout and Presentation

Issues related to how a test appears on the page or screen include such details as text and graphic placement, font size, line length, contrast, blank space, splitting words across lines, splitting words or paragraphs across pages, scrolling, and so forth. Test developers know the conventions and constraints in their tests and have them in mind when reviewing the formatting of questions and the layout of tests. Here are two examples of types of material that may have to be modified or revised, to the extent reasonably possible, because they are not visually accessible:

1. Some readers may miss text that is indented or centered. Most text should be flush left since “ragged–right” margins are easier to read. However, for programs that require the use of authentic text, such as ads or train schedules, flush-left style can be a problem. Length and format will be determined by the material itself.
2. Some test takers using electronic modes of delivery have trouble keeping track of what is on each part of a split screen. The need to scroll information also can make it difficult for a test taker to follow the material. Using smaller-than-recommended fonts may reduce the need for scrolling but may make the material difficult to read. It may be possible to arrange the material so that meaningful units of text or graphics can be viewed together.

## **Section 4: Development of Accessible Nontest Materials**

Accessibility is a concern not only for test materials but for other test-related products and services as well. These include practice materials, program bulletins, score reports, Web sites, and other methods of communication with test takers and schools. Accessibility strategies for nontest materials will vary depending, among other things, on the purpose of the material in question and its anticipated audience. The information in this section pertains largely to nontest materials that accompany postsecondary tests at ETS. In the K–12 setting, practice materials are provided in many different ways. Nevertheless, in K–12 as well, ETS strives to provide such materials.

### **Practice Material**

ETS produces practice tests and test questions in the various alternate formats used by individuals who are blind or have low vision. For blind or low-vision test takers, practicing with the actual format can be as important as practicing the content. An online practice test may be sufficient preparation from a content standpoint, but a person who is blind may also want to know how the test will be laid out, what kind of tactile graphics will be provided, whether any unusual transcription conventions must be learned, and so on. As online practice materials become more common, the general issues of Web accessibility for people with visual disabilities are of increasing importance, in addition to the specific issue of whether the Web version will accurately represent the test to an individual using an alternate format.

### **Familiarization Material**

Test familiarization material is intended to provide users with general information about a test—for example, its length and the topics and types of questions included—rather than actual sample questions. A fully accessible online version of such material is usually provided for blind test takers, who typically access test content via a screen reader. However, if any of that material is inherently graphical and the screen-reader software cannot read it meaningfully, tactile graphics (to supplement the Web version) or an alternate format of the familiarization material may generally be provided if it is reasonably possible to do so.

## **Web Sites**

As more and more content becomes available online, Web access becomes increasingly important. Some recent court cases have held that the Internet is a place of public accommodation and hence is subject to the Americans with Disabilities Act (ADA). ETS strives to make the ets.org Web site as accessible as reasonably possible to users with disabilities.

Online practice materials require special attention from assessment staff knowledgeable about the validity of test content. For example, even descriptions of graphics that are legally sufficient may be inadequate for test takers with certain disabilities. Also, descriptions of graphics in online test-preparation materials should follow the conventions used in the actual test to ensure that the practice materials serve their intended purpose.

## Appendix A: Overview of Test Adaptations

This section provides basic information that describes test adaptations typically used for individuals who are blind or have low vision (braille, tactile graphics, large print, audio), for individuals who are deaf or hard of hearing, and for individuals with learning disabilities.

Types of testing accommodations can be organized into six categories: presentation, response, timing, scheduling, setting, and other. Individuals who adapt tests have to be concerned with the first three—presentation, response, and (occasionally) timing accommodations. The material in this section focuses on presentation issues but notes response or timing issues where relevant.

### Background

When adapting tests into various alternate formats, rendering the material literally and exactly as in the original print or on-screen representation must sometimes be secondary to conveying the intended meaning and points to be tested, in order to maximize comparability of scores.<sup>22</sup> In fact, those who adapt tests may need to add or remove information. For example, they generally add text to scripts for recorded or live audio, but they remove illustrations and student misspellings that are included for the appearance of authenticity in standard editions of a test, as well as map markings or text-formatting details not relevant to the particular testing objective.

In making decisions about what should be added or removed, staff who are creating the adaptations should seek informed professional judgment about the testing role of the question components involved. For example, several rivers and streams labeled in very small print may be displayed on a map. Because text on a figure in a large-print test is in the standard large-print size, larger labels will run together or overlap. If the names of some of the rivers and streams are not relevant to the question in any way (e.g., needed as background information, for keying, or as distracter fodder), then the labels and perhaps the rivers and streams should be removed. A test

---

<sup>22</sup> Many layout features are incidental and need not always be duplicated in alternate formats. For example, double-column text may be converted to single-column text, italics converted to underlining, and various changes made to graphics. Sometimes changes to text may also be helpful (where construct appropriate), for example, to better integrate the question's stem with the description of its graphic. Overliteral interpretations, particularly of incidental layout features or common print conventions, can create confusion.

developer with a thorough understanding of the construct and of how the question is meant to function should decide which elements can or should be removed.

In a language test in which the construct is typically more literal than conceptual, small wording changes may make a significant difference in the performance of the question. However, it still may be necessary and appropriate for accessibility purposes to make display-format changes or to convert images of realistic text (e.g., train schedules) into actual text.

## **Adaptations Usually Approved for Individuals Who Are Blind or Have Low Vision**

Note that each of the adaptations described below is generally approved along with some amount of extended time.

### **Braille**

Braille is a system of representing print symbols by the use of raised dots. For example, the word “braille” in uncontracted braille looks something like this:

⠠⠠⠠⠠⠠⠠⠠

There are several varieties of braille; ETS follows the standards of the Braille Authority of North America (BANA), available at <http://www.brailleauthority.org/>. The most common braille types are:

- uncontracted—one symbol used for each printed symbol; most often used for early elementary grades or foreign languages; alphabetic in that one symbol represents each letter of the alphabet, and there are additional symbols for punctuation marks, etc.; may occasionally be approved for a foreign test taker not fully fluent in English braille code
- contracted—single symbols used for some common words or common letter combinations; e.g., a single symbol is used for “the” rather than spelling the word out
- Nemeth code—used for mathematics and science; uses the same dot combinations as standard contracted literary braille but with different meanings to represent numbers and mathematical symbols



Most of the tests at ETS are brailled using contracted braille and Nemeth code for mathematics and science tests. For various reasons, braille foreign languages, especially non-Western languages, can be challenging, as is braille pieces of text formatted in a nonlinear way—sidebars, flowcharts, some poetry, and proofreading marks superimposed on text. In addition, special symbols, music, and special fonts and symbols used in computer science material can be challenging to braille. A braille expert should be consulted early in any project involving such complications in order to assure that decisions are made with due consideration of how braille practices intersect with testing issues. Often, a proofreading step is included after the braille version is complete, when a braille expert reads the test form to a test developer for approval.

### **Tactile Graphics**

Graphics often can be adapted and made usable for blind and low-vision test takers. Graphics that are adapted for use by touch are called tactile or raised line graphics; labels on such graphics are done in braille. The usability of a graphic depends on its role in the test, the amount and type of detail, and whether the graphic is in a standard format appropriate to the subject matter. Bar charts, line graphs, circle graphs, etc., all have more or less standard formats and components. A blind individual is more likely to have encountered tactile versions of these. A photo, sketch, or representation of a toddler's scribble is nonstandard, may not have been encountered before, and is next to impossible to interpret by touch. Others fall in the middle: it might be possible to adequately simplify diagrams of cellular structures when the diagrams are given along with appropriate text descriptions ("Slide one shows three sets of chromosomes . . .").<sup>23</sup> Parts of the tactile graphics might also be given one-word or very brief descriptions ("burner," "candle," "wing," "webbed foot," "dark clouds," etc.) if they are essential aspects.

Because individuals differ in their skill at reading tactile graphics, a verbal description is also needed to guide the test taker. This description should be thorough but as concise as possible, should consider grade level and vocabulary level, and should convey needed information without compromising testing of the construct.<sup>24</sup> For a braille test, the tactile

---

<sup>23</sup> The American Printing House for the Blind (APH) has provided guidelines for creating tactile graphics. Both APH and the American Foundation for the Blind are resources for standardized symbols and tactile features on maps.

<sup>24</sup> Previously, it was thought that a very brief description would be adequate (e.g., replacing "the figure above" with "the square"). Since then, however, research and consultation with blind and sighted experts on accessibility indicate that even a skilled tactile-graphics user is

graphics are generally included in their correct position in the test itself, usually on a facing page (for technical reasons they often cannot be included on the same page as braille text). For some tests or practice materials, it might be preferable to produce the tactile graphics in a separate booklet and to refer to them in the braille text. That may be appropriate for cases in which the same graphic or related graphics are to be used with different pages of the text. Tactile graphics can also be put into a separate booklet so audio users who need them can use them. Some test takers may benefit from tactile graphics that are visually enhanced; it may not be feasible, however, to produce these in addition to the standard kind.

Not every blind individual can read braille well<sup>25</sup> or prefers it over other media such as audio. For example, individuals who have become blind later in life or who have physical impairments that limit their ability to read tactile information are likely to rely on audio. Braille may not be the primary or only adaptation needed for blind test takers.

## **Audio**

Audio delivery is often used by test takers who are blind as well as by test takers with learning disabilities (see below for a brief discussion of some differences in the ways these groups typically use audio). Audio takes a variety of forms: prerecorded speech provided via cassette or CD, live reader, or text delivered by synthetic speech (including screen readers such as JAWS<sup>®</sup> or a self-voiced system such as that used in the Voiced GRE<sup>®</sup>, where the computer-generated audio output is integrated into the test software).

---

likely to need the more detailed description. In the example of a square, we would add information about any labeled vertices, intersecting lines, etc. Essentially, descriptions accompanying tactile graphics guide the user through the graphic as much as possible, without overwhelming him or her with excess information. It can help to begin a description in general terms and then fill in details. Most users will follow along with a tactile graphic, but some may rely on the description alone. Complicated graphics that cannot be described concisely (including representations of three-dimensional figures) may have to be dropped.

<sup>25</sup> One commonly reported statistic is that approximately 11 percent of blind individuals read braille, but the proportion depends on a person's age, the age at onset of blindness, and which individuals were considered "blind" for the purposes of gathering the data.

Most tests being prepared for recorded audio delivery require scripting, markup, or other extensive preparation. A detailed script is prepared that specifies pronunciation of any difficult words or proper names, words that must be spelled aloud, instructions for dealing with blanks or underlined sentence fragments, precise wording of figure descriptions or verbalizations of math expressions, and the like. In addition, those adapting tests for audio delivery may provide large-print or tactile figure supplements if the test includes figures.

For postsecondary programs in which ETS determines the parameters for the use of a live reader as an accommodation, we generally find it necessary to provide a reader script, very similar to a recording script, as well as tactile or enlarged graphics as needed, for blind or low-vision test takers. (Because sighted test takers with learning disabilities can take in visual elements of text and/or graphics, scripting is less critical for this population and may even be confusing.) We find that readers who describe visuals on their own for a blind test taker may provide information that is excessive, insufficient, inconsistent, or inappropriate. The few exceptions that do not call for reader scripts or notes are tests that consist only of straightforward text, without such complications as difficult pronunciations, words to be spelled out, mathematical or scientific formulae, visuals/graphics, special text formatting, musical notation, etc. For a self-voicing test such as Voiced GRE, scripts are based on recordings or reader scripts but need adjustments and additional markup to guide pronunciation and navigation.

The most crucial difference between audio delivery and print or braille delivery is that audio is of necessity a linear format: for the test taker, the content comes and goes and cannot be quickly scanned or reviewed for crucial information in the same way that print or braille can. As a result, it is essential to consider memory load both in determining whether it is appropriate to adapt a question and in producing the adapted version, if the adaptation itself is appropriate. How much information must the listener retain in memory to make sense of the adapted version of the question? Is it so much greater than the memory load required for the standard version as to make the question essentially inaccessible in audio? Consider the following scenarios:

- Reading passages—a print or braille user can easily go back to a long reading passage to find information to answer a test question. An audio user, however, must be provided with ways to easily navigate the material without having to listen to all of it again (unless finding

the information is an essential aspect of the test). Line numbering is useful in print or braille but not in audio.

- Questions with figures (including graphs, charts, maps, or illustrations) that must be described—a test taker may expend more effort in trying to remember a lengthy description than in working through the question, and the function of the question may be seriously undermined.
- Chunks of text containing blanks that the test taker must fill in (e.g., GRE sentence-completion and text-completion questions, various language tests) or that contain labeled portions from which the test taker is to identify a portion with an error—the longer the text, the more blanks included, and the more options for filling in each blank, the greater the potential difficulty.

Navigation aids can reduce the problem of memory load. The types of aids available vary with the type of audio and can include the ability to skip parts of a question, tone indexing for cassettes and CDs, track management for CDs, instructions for live readers, or hyperlinks for materials provided electronically. (Note, however, that in certain assessments such as those for foreign languages, some of these aids can affect what is tested.)

Materials that typically need special attention when being prepared for audio delivery include:

- graphics
- mathematics or other notation
- certain conventions or formats (e.g., underlining in grammar-usage questions, questions in a Roman-numeral format in which answer choices include "I only," "I and II only," etc.)
- pronunciation of key words (noted in the script unless determining correct pronunciation is part of the construct)
- spelling of some words (especially homophones) in scripts
- spelling and phonetics or phonics questions
- questions employing visual strategies
- depictions of handwritten material
- punctuation that should be read aloud
- response directions (especially if responses will be entered by a scribe)

## **Large-Print Tests and Figure Supplements**

Large-print tests are generally used by test takers with low vision and occasionally by those with learning or physical disabilities. Large-print supplements may accompany audio formats for low-vision individuals. Staff who help to prepare large-print tests should familiarize themselves with the requirements of the tests they work on so that they can communicate appropriately with formatters and layout specialists.

One format that is distinct from “large print” is “enlarged type,” which generally uses a 14-point font size and is created through a photo-enlargement process. The process is relatively straightforward and generally does not involve rearranging or modifying elements of the test material.<sup>26</sup> Enlarged-type formats are either done at the test taker’s request or are specified by certain programs.

Large-print materials, on the other hand, usually have to be reformatted in the larger size and thus require more attention from assessment staff. “Large print” refers to print that is 18 points or greater (for reference, the body text in this document is 12 points). Some testing programs that make large-print editions in advance (rather than on demand) use 18-point text; others use 20 points or larger.<sup>27</sup>

When questions are reformatted for large-print tests, relative placement of questions and stimuli may shift, affecting references to figures and other information (e.g., “above,” “below,” or “on page x”). Line breaks will also change, and so all references to line numbers may need to be checked and adjusted. Graphics, including musical scores, may need to be turned to landscape mode to accommodate their size and dimensions, and the document may need to be bound so that the pages can be folded out.<sup>28</sup>

---

<sup>26</sup> However, in the event that questions on a photoenlarged test require test takers to measure the length of lines, a two-inch line will no longer be two inches if enlarged.

<sup>27</sup> New programs that intend to produce large-print tests must decide what font size to use. At ETS, requests for 24 point are relatively rare, and many test takers find 24-point type too large. Also, the larger the point size, the more complications there are in reformatting. ETS recommends 18 or 20 point as the standard size for large-print tests, with larger type for multilevel headings. This follows the practice of the American Printing House for the Blind for its Accessible Textbooks ([http://www.aph.org/atic/atic\\_lp\\_form.html](http://www.aph.org/atic/atic_lp_form.html)).

<sup>28</sup> Foldouts, however, are kept to a minimum; they can confuse test takers or be difficult to handle physically.

If a test that includes graphics is produced in an audio format, a large-print figure supplement is generally also produced for postsecondary tests. In such a supplement, all labels on the graphics are at least 18 point or larger, and the graphics are enlarged at approximately the same ratio as the print. In measurement questions, graphics often cannot be enlarged (e.g., a ruler intended to be life-size but enlarged to double that size will usually not serve the intended purpose).

Text that usually appears in smaller print in a regular format (such as footnotes or superscripts) will not be large enough for a large-print user to read. Generally, ETS uses the same size that is required for the larger point size or one slightly smaller (e.g., a 20-point test might have 18-point superscripts and subscripts). In addition to font size, other aspects of a test may need to be adapted for the large-print edition:

- Typefaces—These should be clear. Verdana is generally best for regular text.<sup>29</sup> It provides slightly better differentiation between some characters (e.g., “l” and “1” are not distinguishable in Arial but they are in Verdana). A serif font such as Times or Times New Roman provides well-differentiated italics for math questions.

### This is Verdana typeface in 18-point text.

- Font style—Emphasis is best conveyed by bold, underlined text (not italics). If possible, long stretches of italics should be avoided. Decorative or script typefaces and type that is not horizontal (e.g., in illustrations) should also be avoided. Words printed in all-capital letters are also difficult to read, so title case is preferred for titles of figures or graphs.
- Graphics—These can be made more accessible by adjusting the contrast and by reducing visual clutter.
- Margins—Paragraphs should be flush left with at least one-inch margins. It is helpful to insert a space between paragraphs or to provide a small first-line indent to indicate paragraph boundaries. In general, indented copy should be kept at a minimum when reasonably possible.
- Spacing—Spacing between lines, often referred to as “leading,” should be at least 1.25 spaces.

---

<sup>29</sup> <http://www.aph.org/advisory/slate/spring2001/slatesp01.pdf>, p. 13.

These and additional practical guidelines are available from the American Printing House for the Blind at <http://www.aph.org/>.

## **Adaptations Usually Approved for Deaf or Hard of Hearing Individuals**

Some common adaptations approved for deaf or hard of hearing individuals are listed below. The use of each accommodation for a particular test must be considered in line with how the accommodation interacts with the construct being measured and with how best to maximize score comparability with tests given without accommodations:

- omitting, allowing for replay, or providing written transcripts<sup>30</sup> of listening sections (listening sections are not limited to language tests)
- omitting or adapting speaking sections or permitting written responses
- omitting sound-related content, such as music, rhyme, or phonics, if construct appropriate
- scripting a listening section for an oral interpreter

Extended time is sometimes approved in combination with these adaptations. In some cases a test taker with a hearing loss requests extended time to permit replaying of audio selections. The permissibility of this accommodation will depend on test construct. As mentioned elsewhere, if replaying audio content (beyond what the standard test permits) is permissible on construct grounds, it is extremely helpful if the audio material is designed to facilitate this.

ETS does not recommend on-the-fly interpretation of test questions or test directions into American Sign Language (ASL) or other sign languages or sign codes as an accommodation. These accommodations may significantly alter test construct and interfere with standardization. In some programs in which the client, state, or school is responsible for determining reasonable accommodations, on-the-fly interpretation may be permitted.<sup>31</sup>

---

<sup>30</sup> Transcripts will change the construct in some foreign-language tests as well as in some ETS professional-licensure tests.

<sup>31</sup> ETS recommends translations of high-stakes tests into foreign languages only if translation is deemed compatible with the test construct and if these translations are created, recorded, back-translated, and validated in advance, in accordance with research-based best practices for test translation. ASL is a language in its own right, distinct from English or any other spoken language; therefore, the same principles apply as for any other

## **Adaptations for Test Takers with Learning Disabilities**

The most common accommodations for individuals with learning disabilities (assuming no additional sensory or motor disability) are extended time, extra breaks, and/or delivery by a live reader. Such individuals do not typically need any test adaptations.

### **Extended Time**

Many K–12 tests are untimed to begin with. In most of the programs that have timed tests, extended-time accommodations do not call for test adaptations (only modifications to references to timing in the directions) and so do not require significant attention from test development staff. There are exceptions. For example, some tests with listening sections use audio that may need to be paused, repeated, or rerecorded with longer pauses for responding. Changes may have to be made to the audio recording and/or to the test directions to facilitate navigation.

### **Live Reader and Other Accommodations**

It is often acceptable for a live reader to read the actual test to a test taker with a learning disability, without using a script. Sometimes a tip sheet can give the reader enough information to do so. Unlike a test taker who is blind, a test taker with a learning disability is usually able to take in the visual elements of the test and/or graphics. However, large-print tests or reader scripts may be needed for test takers with learning disabilities in certain cases. Some test takers with learning disabilities are approved for recorded audio, in which case scripts and recordings can be produced as they would be for a blind test taker.

---

language. Signing is permitted for spoken directions in the test center, such as those used when showing test takers to their seats or checking identification.



## **Accommodations and Adaptations for English Language Learners**

Although being an English Language Learner (ELL) does not mean a student has a disability, ELLs represent a subpopulation for which accommodations and/or test adaptations are sometimes appropriate. For a full discussion of issues related to assessing ELLs, see the publication *ETS Guidelines for the Assessment of English Language Learners*. It should be noted that some students with disabilities are also ELLs; decisions about accommodations and test adaptations for those students take both characteristics into account.

### **Use of Technology**

Various forms of technology, including calculators (scientific, graphing, talking, with software, etc.), are sometimes required or permitted in tests for all students and sometimes prohibited, depending on the test construct. In some cases, a permitted or required technology (such as a graphing calculator) may be inaccessible to test takers with certain disabilities. In such cases, it may be possible to provide accessible alternatives or to make adaptations to the test. Sometimes calculators are approved as an accommodation on tests that otherwise prohibit them. Similar issues can arise with other forms of assistive technology, such as voice-recognition software, spell-checkers, etc.

## Appendix B: References on Accessibility

- Allman, C. 2004. *Making Tests Accessible for Students with Visual Impairments: A Guide for Test Publishers, Test Developers, and State Assessment Personnel*. (2nd edition.) Louisville, Ky.: American Printing House for the Blind (APH). <http://www.aph.org/>
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 1999. *Standards for Educational and Psychological Testing*. Washington, D.C.: AERA, APA, and NCME.
- Amick, N. et al. 1997. "Guidelines for Design of Tactile Graphics." APH Educational Research. <http://www.aph.org/edresearch/guides.htm>
- Arditi, A. 2008. "Making Text Legible—Designing for People with Partial Sight." Lighthouse International. <http://www.lighthouse.org/accessibility/design/accessible-print-design/making-text-legible/>
- Dolan, R. P., and T. E. Hall. 2001. "Universal Design for Learning: Implications for Large-Scale Assessment." *IDA Perspectives* 27 (4): 22–25.
- "Effective Practices for Description of Science Content Within Digital Talking Books." Carl and Ruth Shapiro Family National Center for Accessible Media. <http://ncam.wgbh.org/dtb/>
- ETS Guidelines for Fairness Review of Assessments*. 2009. Princeton, N.J.: Educational Testing Service.
- ETS Guidelines for Fairness Review of Communications*. 2009. Princeton, N.J.: Educational Testing Service.
- ETS Guidelines for the Assessment of English Language Learners*. 2009. Princeton, N.J.: Educational Testing Service.
- ETS International Principles for Fairness Review of Assessments*. 2007. Princeton, N.J.: Educational Testing Service.
- ETS Standards for Quality and Fairness*. 2002. Princeton, N.J.: Educational Testing Service.

- Hansen, E. G. et al. 2005. "Accessibility of Tests for Individuals with Disabilities Within a Validity Framework." *System: An International Journal of Educational Technology and Applied Linguistics*, 33(1): 107–133.
- Hansen, E. G. et al. 2008. *Evidence Centered Assessment Design for Reasoning About Testing Accommodations in NAEP Reading and Mathematics (ETS RR-08-28)*. Princeton, N.J.: ETS Research Report Series.
- Hansen, E. G., and R. J. Mislevy. 2006. "Accessibility of Computer-Based Testing for Individuals with Disabilities and English Language Learners Within a Validity Framework." In *Online Assessment and Measurement: Foundations and Challenges*, edited by M. Hricko and S. Howell, 214–262. Hershey, Pa.: Idea Group Publishing, Inc.
- Johnstone, C. J. et al. 2008. "Universal Design and Multimethod Approaches to Item Review." *Educational Measurement: Issues and Practice* (Spring). Madison, Wis.: National Council on Measurement in Education.
- Ketterlin-Geller, L. R. 2005. "Knowing What All Students Know: Procedures for Developing Universal Design for Assessment." *Journal of Technology, Learning, and Assessment* 4 (2).  
<http://escholarship.bc.edu/jtla/vol4/2/>
- Kitchel, J. E. 2004. "Large Print: Guidelines for Optimal Readability and APHont™ a Font for Low Vision." American Printing House for the Blind. <http://www.aph.org/edresearch/lpguide.htm>
- Laitusis, C. C. 2007. "Validity and Accommodations: A Variety of Approaches to Accessible Assessments." In *Large-Scale Assessments and Accommodations: What Works?*, edited by C. C. Laitusis and L. L. Cook. Washington, D.C.: Council for Exceptional Children.
- National Accessible Reading Assessment Projects (NARAP). 2006. "Defining Reading Proficiency for Accessible Large-Scale Assessments: Some Guiding Principles and Issues."  
<http://www.narap.info/publications/reports/definingreadingprof.htm>
- National Center on Educational Outcomes (NCEO) Online Accommodations Bibliography. <http://apps.cehd.umn.edu/nceo/accommodations/>

National Research Council, Committee on Participation of English Language Learners and Students with Disabilities in NAEP and Other Large-Scale Assessments. 2004. *Keeping Score for All: The Effects of Inclusion and Accommodation Policies on Large-scale Educational Assessment*, edited by J. A. Koenig and L. F. Bachman. Washington, D.C.: National Academies Press.

Thompson, S. J., C. J. Johnstone, and M. L. Thurlow. 2002. "Universal Design Applied to Large Scale Assessments." Minneapolis: University of Minnesota, National Center on Educational Outcomes.  
<http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>

Thurlow, M.L. et al. 2009. "Accessibility Principles for Reading Assessments." Minneapolis: National Accessible Reading Assessment Projects.  
<http://www.narap.info/publications/reports/narapprinciples.htm>

Wendler, C., and D. Powers. 2009. *What Does It Mean to Repurpose a Test?* Princeton, N.J.: Educational Testing Service, R&D Connections #9.

## Appendix C: Glossary

**Accessibility**—The extent to which all intended test takers with disabilities are able to obtain a valid score that is a valid measure of the construct being measured on the test regardless of the test takers' disabilities. Accessibility is often enhanced by accommodations. See **Accommodations**.

**Accommodations**—A change to a test, its administration site, its timing, its presentation mode, and/or its response mode to allow access to the test for a person with a disability. In some usages, an **accommodation** refers to a change that does not alter the knowledge, skill, or other attribute that the test is intended to measure; the word **modification** is then used for a change that does alter the knowledge, skill, or other attribute that the test is intended to measure. See **Accessibility, Presentation Mode, Response Mode**.

**ADA (Americans with Disabilities Act)**—Civil rights legislation in the United States that addresses discrimination against persons with disabilities in the public and private sectors, including all public services, public accommodations, transportation, and telecommunications. Section 12189 states: "Any person that offers examinations or courses related to applications, licensing, certification, credentialing, for secondary or postsecondary education, professional, or trade purposes shall offer such examinations or courses in a place and manner accessible to persons with disabilities or offer alternative accessible arrangements for such individuals."

**Alternate Test Format (ATF)**—Test prepared in any nonstandard format for delivery to individuals with disabilities who cannot use the standard format. Examples include braille, large print, reader script, recorded audio, and computer-voiced. Alternate formats can also include tests from which sections (e.g., a listening section) have been removed for purposes of disability accommodation.

**AT Act (Assistive Technology Act)**—1998 federal legislation that provides financial assistance to states for maintaining and strengthening statewide programs of technology-related assistance for individuals of all ages with disabilities.

**Blindness, Legal Blindness, Low Vision**—Definitions vary. For one useful set, go to: <http://www.lighthouse-sf.org/services/visionlossfaq.php>.

**Construct**—All of the knowledge, skills, or abilities (KSAs) a test is intended to measure with minimal influence from other KSAs that the test is not designed to measure. For example, a test intended to measure the construct of quantitative reasoning may also inadvertently measure verbal reasoning as well if the questions are unnecessarily difficult to read. See **KSA**.

**Construct-Irrelevant Score Variance**—Differences in scores among test takers attributable to knowledge, skills, or abilities (KSAs) that the test is not intended to measure. For example, scores on a quantitative reasoning test may vary if the questions are unnecessarily difficult to read, because some test takers are better readers than others. See **Construct, KSA, Variance**.

**Construct Underrepresentation**—Construct underrepresentation occurs when a test either implicitly or explicitly purports to measure a particular construct and does not test significant aspects of the construct. For example, a test may purport to measure "communicative ability" yet actually measure only reading and writing, even though speaking and listening are also important aspects of the construct of communicative ability; or a test that is supposed to measure the outcomes of a college-level chemistry course may not measure laboratory skills, even though such skills are an important aspect of the construct of college-level chemistry.

**Decoding**—The part of the reading process involved in translating marks on a page to letters and words. If decoding is part of the intended construct, then reading tested material to the individual as an accommodation for a disability changes the construct being tested. If the construct is comprehension rather than decoding, reading tested material to the individual as an accommodation for a disability would leave the construct relatively unchanged. See **Accommodation, Construct**.

**Disability**—A physical or mental impairment that substantially limits one or more major life activities. The Americans with Disabilities Amendments Act, which took effect January 1, 2009, as Public Law 110-325, provides a specific list of activities and bodily functions that are to be considered "major life activities" and notes that the list is not intended to be exhaustive: <http://www.access-board.gov/about/laws/ada-amendments.htm>.

**Essential**—In the context of revising test questions to increase accessibility, an essential aspect of the question is one that cannot be eliminated without changing the construct measured by the question. See **Accessibility, Construct, Helpful, Incidental**.

**Flagging**—Including an indicator, such as an asterisk, on a score report to inform score users that a reported score was obtained under nonstandard conditions that may make the reported score difficult to compare to scores obtained under standard conditions.

**Helpful**—In the context of revising test questions to increase accessibility, a helpful aspect of the question is one that improves the test’s ability to measure the intended construct but is not essential because the test can still measure the construct without it. See **Accessibility, Construct, Essential, Incidental**.

**IDEA (Individuals with Disabilities Education Act)**—Provides for services to children with disabilities throughout the United States.

**Incidental**—In the context of revising test questions to increase accessibility, an incidental aspect of the question is one that can be eliminated without changing the ability of the question to measure the intended construct and without incurring significant loss to the quality of the question. See **Accessibility, Construct, Essential, Helpful**.

**JAWS**—Currently the most commonly used screen-reader software. It works with Microsoft Windows. More information is available at <http://www.freedomscientific.com/jaws-hq.asp>.

**Key**—The answer to a multiple-choice test question.

**KSA**—Knowledge, skill, or ability; in some usages, knowledge, skill, or other attribute. **KSA** is often used to stand for all of the attributes of test takers that a test is intended to measure.

**Leading**—The vertical space between lines of print.

**Learning Disability**—A disorder manifested by significant difficulty in listening, speaking, reading, writing, reasoning, or doing mathematics. These disorders are intrinsic to the individual and presumed to be due to central nervous system dysfunction rather than sensory impairments, developmental disabilities, or cultural differences. Learning disabilities often occur with other disabilities such as attention-deficit/hyperactivity disorder (ADHD).

**Legal Blindness, Low Vision**—See **Blindness**.

**Modification**—See **Accommodation**.

**Office of Disability Policy (ODP)**—Part of the ETS Office of Professional Standards Compliance, which is in turn part of the Office of the General Counsel. The ODP establishes and disseminates ETS policies about accessibility and accommodations for individuals with disabilities who use ETS products and services. The ODP also oversees the work of an external panel of accommodations experts who review requests for accommodations from individual test takers. The assessment area evaluates the appropriateness of an accommodation relative to test construct.

**Plain English**—A communication style that focuses on the needs of the audience and avoids unnecessary complexity. An assessment is in Plain English if the test questions and other test materials are presented in clear, accessible language that is consistent with valid measurement of the construct. This can be achieved by integrating Plain English reviews into the development process for test forms to be administered to all students. Plain English reviews undertaken after the test is developed may be conducted to develop distinct, accommodated Plain English test forms; these accommodated forms are typically administered to test takers who are English Language Learners, who have reading difficulties, or who have other special needs related to language.

**Population**—All of the members of some defined group, such as third-grade students in the United States. Many populations are too large for all members of the population to be tested. In those cases, smaller samples of members are drawn from the population. The population is the total group that a sample is intended to represent. Compare **Sample**.

**Presentation Modes**—The means by which test questions are administered to test takers, including, but not limited to, printed booklets, large-type printed booklets, braille booklets, computer screens, computer-voicing, and human voice (live or recorded). Compare **Response Mode**.

**Pretest (Field Test, Pilot Test)**—A collection of test questions administered to obtain information about the characteristics of the questions before the questions are used to obtain information about test takers.

**Read Aloud**—Any accommodation in which the test is read out loud to the test taker. The term “read aloud” is the usual term in K–12 testing; “reader,” “reader delivery,” or “live reader” is often used for other tests. A student who qualifies for “read aloud” might receive either a live reader or a recording as an accommodation.



**Reader Script**—A script provided to a reader in order to ensure that test content is delivered to a test taker exactly as intended. It may include figure descriptions, pronunciation indicators, and/or words to be used to describe mathematical or scientific expressions or other special symbols, plus any other necessary instructions to the reader. Not all testing organizations provide reader scripts, and not all tests require reader scripts for valid administration, but they are recommended whenever a reader would be expected to need guidance in order to ensure validity.

**Rehabilitation Act of 1973**—Federal legislation that provides, in Section 504, that: “No otherwise qualified individual with a disability in the United States . . . shall, solely by reason of her or his disability, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance.”

**Response Modes**—The methods by which test takers indicate their answers to test questions, such as filling out an answer sheet, writing an essay, speaking, clicking a mouse, or performing some task such as parallel parking a car. Compare **Presentation Mode**.

**Rich Text Format (RTF)**—A document format that retains significant formatting information and can be imported and exported by a large number of computer word-processing and layout programs.

**Sample**—A group (often, but not necessarily, a group of people) intended to represent a larger population. For example, a group of individuals taking a pretest may be intended to represent the eventual test-taker population. Compare **Population**.

**Screen Reader**—Computer software that converts information on the computer’s screen to synthesized speech and allows the keyboard to take the place of a mouse for most or all interactions with the computer. Screen readers are primarily intended for use by individuals who are blind or have low vision. See **JAWS**, **Text Reader**.

**Section 508**—Part of the 1998 amendments to the Rehabilitation Act of 1973. Section 508 provides minimum technical standards for accessibility that technology must satisfy. Section 508 is a separate statute from the Americans with Disabilities Act.

**Self-Voiced**—In a self-voicing system, computer-generated audio output is integrated into the test software itself, rather than delivered via off-the-shelf

screen-reading (voicing) software such as JAWS or Window-Eyes. Self-voicing permits ETS to control exactly how each question is read and allows blind and low-vision test takers to navigate the test and individual questions using the keyboard and voiced prompts. See **Voiced GRE**.

**Stem**—The initial portion of a multiple-choice question that does not include the answer options.

**Stimulus Materials**—Information given to a test taker on which one or several test questions are based. Common stimulus materials include reading passages, graphs, tables, diagrams, photographs, maps, videos, and speech, such as a portion of a lecture.

**Tactile Graphics**—Figures, such as maps, graphs, and geometric figures, created in raised relief from the page so that the lines can be felt by a blind individual. Any required labels for tactile graphics are typically provided in braille for the benefit of those users who can read braille.

**Test Adaptation**—Modifications made to a test to create alternate test formats. See Appendix A for a full discussion.

**Test Familiarization**—Preparing to take a test by learning about the characteristics of the test. Test familiarization materials may include descriptions of the content and KSAs to be measured (in general terms, without revealing the specific questions on the operational test), the numbers and types of questions in the test (e.g., multiple choice, essay), the amount of time allowed, the response mode(s), and so forth. Familiarization materials often include sample test questions. As the number of sample questions increases, test familiarization becomes similar to test practice. See **Response Modes**. Compare **Test Practice**.

**Test Practice**—Preparing to take a test by responding to a test similar to (but **not** identical to) the edition of the test that will be used operationally. If the practice test questions are too similar to the actual test questions, test practice can result in inappropriately inflated scores. Compare **Test Familiarization**.

**Text Reader**—Software that uses text-to-speech to read selected text to a user. Unlike a screen reader, a text reader does not attempt to render everything on the screen and may require using a mouse to select the text to be read. Text readers are commonly used by individuals with learning disabilities and some with low vision. See **Screen Reader**.

**Universal Design**—A concept originating in architecture, universal design promotes designs that increase access without post-hoc adaptation (for example, designing a building to have ramps as well as steps). In the context of testing, it is an attempt to make test materials accessible to a wide range of test takers, including people with disabilities, by eliminating unnecessary barriers to access, such as language that is more difficult or visual materials that are more complicated than are required for valid measurement. Tests that are universally designed may still require adaptation to accommodate certain disabilities. See **Accessibility**, **Validity**.

**Validity**—The extent to which inferences and actions made on the basis of a set of scores are appropriate and justified by evidence; the extent to which a test is meeting the purpose for which it was intended.

**Variance**—Differences among test scores. Technically, variance is a statistic indicating the average squared difference between each test score and the mean test score.

**Voiced GRE**—A self-voicing computer-delivered version of the GRE General test. This version is intended for blind and low-vision test takers. See **Self-Voiced**.