Training Module for the Universal Design of English Proficiency and Content Area Assessments

In collaboration with:
Council of Chief State School Officers (CCSSO)
National Association of State Directors of Special Education (NASDSE)

Supported by:
U.S. Office of Special Education Programs
Training Module for the Universal Design of English Proficiency and Content Area Assessments

These materials were created for the Minnesota Department of Education by Sandra Thompson, Michael Anderson, and Kristin Kline Liu

2004

Included in this module:

- Presentation slides and notes for universal design training
- Considerations for universal design of assessments, references, and universal design worksheets for item review
- Examples of revised test items for use in training on using the considerations

All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as:

This training manual was funded, in part, by a 2003 subgrant from the Minnesota Department of Education as part of the U.S. Department of Education Title VI grant (CFDA84.368A): Improving the Achievement of English Language learners on Statewide Proficiency Tests through Authentic Proficiency Assessments. The grant was awarded to the Minnesota, South Carolina and Nevada Departments of Education. It was also supported by the National Center on Educational Outcomes (NCEO). NCEO is supported through a Cooperative Agreement (#H326G05007) with the Research to Practice Division, Office of Special Education Programs, U.S. Department of Education. Opinions expressed herein do not necessarily reflect those of the U.S. Department of Education or Offices within it.
Universally Designed English Proficiency and Content Area Assessments

Improving the Achievement of English Language Learners through Authentic Proficiency Assessments
May, 2004

This training will address issues of universal design. The ideas and suggestions presented represent a work in progress, a work created through input from many people. As we work together through this training, we will continue to gather the best of all of our experiences in order to build better assessments for all of our students.
We have three main goals for this training:

• The first is to talk about what we mean by “universally designed assessments,” why we need them and who they are for.
• Our second goal is to be able to identify and give examples of “elements” of universally designed assessments. In this part we will also share examples of how these elements are applied with students, using a “think-aloud” technique (also called cognitive labs).
• Finally, we will apply these elements to considerations for reviewing test items.
The goal of universally designed assessments is to provide the most valid assessment possible for the greatest number of students, including English language learners.

Thinking about elements of universal design is important when assessments are designed, and it is important for the continual refinement and improvement of operational assessments.

It has been common in the past to build assessments that are accessible to the majority of students and only think of English language learners after the assessment has been designed.

An assessment can be much more universal (accessible to a wide range of students) if all students are taken into account at the BEGINNING stages of test design.
But, what does that really mean?

• Do we want to change the standard of performance? **NO**
• Can we forget about accommodations if we do this? **NO**
• Is this all figured out – for now and forever? **NO**
• Is this something that will benefit only English language learners? **NO**

1) Universally designed assessments DO NOT change the standard of performance – they are not watered down or made easier for some groups. It makes the assessment more accessible to ALL students so that we have a better measure of a student’s actual skills to ensure the test is measuring what it is supposed to measure.

2) Universally designed assessments are not meant to replace accommodations. Even by incorporating the elements of universal design in assessment design, accommodations may still be needed for some students in the areas of presentation, response, setting, timing, and scheduling. Universally designed assessments should, however, anticipate some of the common accommodations needed and the design of the test should allow for those accommodations to fit into the assessment process.

3) There is a reason we are calling the ideas we will be working with today “considerations.” They are things that should be considered when developing an assessment. They should be talked about openly and decisions should be made, having weighed the pros and cons of different design elements. As the administration of tests changes (e.g. more computer-based testing is used) the universal design considerations will evolve.

4) In addition to English language learners, ALL students benefit from having more accessible tests. Just like we all use curb cuts for uses that the original laws might not have considered, so all students can benefit from a test being more accessible. And, universal design makes tests better for everyone.
Think about universal design in architecture and tool design

- Curb cuts and ramps
- Signage with universal symbols
- Door handles rather than knobs
- Special pen shapes that are easier to hold

- The concept of universal design has been around for a long time. The term has mostly been applied to architecture, but as you can see by this list, it has been applied in many other areas as well.
- Accommodations made for one group have become helpful for all of us.
- These are all things that do not hinder people, but help greatly in making daily life more accessible.
- These design features help all kinds of people in different ways. Think about biking over curb cuts, or automatic doors that open with a push of a button when you are carrying things.
The goal is to provide optimal standard assessment conditions for today’s diverse population of students.

• We want to design assessments with standard conditions that are the best for today’s diverse population of test takers so that we can really tell what students know and are able to do.

• The idea is to incorporate all of these features into STANDARD test design.

• In this way, we will be creating assessments that are more accessible to virtually all students.
Remember this?

OFFICIAL BALLOT, PALM BEACH COUNTY, FLORIDA

Here is the ultimate example of the high stakes consequences of poor design. Remember the 2000 Presidential Ballot from Palm Beach County, Florida?
The design of the ballot was a problem for many voters, especially Hispanic, African American, and elderly voters, resulting in many invalid votes.
Sometimes we get carried away and lose common sense. A lot of what we are talking about today is just common sense, but it is often forgotten.
Assessments Discussed in this Training

• English Language Proficiency Listening/Speaking Test - *Computer Based*
• English Language Proficiency Reading and Writing Test - *Paper/Pencil*
• Grade level Academic Content Area Assessment - *Computer Based*
• Grade level Academic Content Area Assessment - *Paper/Pencil*

• All assessments need to be universally designed.
• In this training, we will focus specifically on how universally designed assessments are important for English language learners, both on tests of English language proficiency, and for grade level academic content area assessments that are used for accountability.
Universally Designed Assessments Can Increase Reliability and Validity of Assessments for English Language Learners

✓ Analyses of large data files (Abedi, 2002) found language as a source of measurement error that can negatively impact reliability of test results for ELLs.
✓ Language has also been found to be a source of construct irrelevance that can have a negative impact on the validity of the results for ELLs (Abedi, 2002).

• Universally designed assessments can increase reliability and validity of assessments for English language learners.
• Jamal Abedi has done some important research in this area. We’ll show several examples of his work later.
• Assessments that are designed without the English language learner in mind are not as valid OR reliable for that population of students.
• English language proficiency exams that are designed specifically for English language learners need to also consider those English language learners with special needs. Think about how these assessments can be more accessible for ALL of the students taking the test.
• Not all English language learners are able bodied, or have perfect vision or hearing.
Students are the end users of assessments – we need to test “usability” with them!

If we are assessing all students, we need to design assessments that consider ALL students and their needs from the BEGINNING of the design process.
Student Characteristics

Just one of many reasons that we need to be talking about universally designed assessments!

• We’re going to do a little exercise to illustrate how important it is to consider the characteristics of students in the development of assessments.
Complete these well-known phrases:

1. A penny saved is . . .
2. An idle mind is . . .
3. Don’t bite the hand that . . .
4. It’s always darkest before . . .
5. Strike while the . . .
6. Two’s company, three’s . . .
7. If at first you don’t succeed, . . .

Have everyone complete these well-known phrases.

•STOP – Please do not turn the page until invited to do so!
Here’s what the kids say!
A penny saved is . . .

. . . not much.
An idle mind is . . .

. . . the best way to relax.
Don’t bite the hand that . . .

. . . looks dirty.
It’s always darkest before . . .

. . . Daylight Savings Time.
Strike while the . . .

. . . bug is close.
Two’s company, three’s . . .

. . . the Musketeers.
If at first you don’t succeed, . . .

. . . get batteries.
What seems obvious and clear to test developers may not be all that obvious and clear to students.

It takes consistent effort and guidance to make sure that test items and tests really are accessible to all students.

• Things that seem obvious and clear to test developers may not be clear to students. That’s why we are focusing on universally designed tests in this training.
This is a BIG DEAL!

Requirements for universally designed assessments are showing up in federal legislation

• Discussions about universally designed assessments are taking place all over the country.
Title I Regulations introduced the need for universally designed assessments –

[Assessments must be] designed to be accessible and valid with respect to the widest possible range of students, including students with disabilities and students with limited English proficiency.

Sec. 200.2(b)(2)

• Though the words “universal design” don’t appear in Title I – the definition is in the regulations
Increasingly, we need to think about students with a variety of different language and cultural background characteristics.

- We need to think about English language learners in the development and operationalization of all types of assessments.
- This includes designing items, piloting items, reviewing items, and getting feedback from students on items. English language learners need to be considered at all of these points.
The number of English Language Learners in schools has increased dramatically

- Represent about 8.4% of all public school students in the U.S.
- Enrolled in almost half of all public schools nationwide

• Important information for test developers!
• This should not be news to anyone, the ELL population is significant throughout the country in every state.
• This is not a subgroup that is diminishing in numbers, it is still growing so we need to plan our assessments with these students in mind.
• In many districts, ELLs (or ELLs and students whose home language is not English) make up a MAJORITY of the students if not close to a majority. Yet still too often these students’ language needs are not taken into account or are considered only in the final stages of assessment development.
• See http://www.ncela.gwu.edu/states for more information on the growing ELL population by state.
These figures come from a recent study done by a company called Development Associates in Arlington, Virginia. They found that ELLs were most likely underrepresented in special education, but given the rapid increase in the total population of ELLs, this group is most likely growing rapidly too. The numbers highlight the importance of thinking carefully about the needs of all the ELLs taking a particular test.


This resource is available through the National Clearinghouse for English Language Acquisition (NCELA) at http://www.ncela.gwu.edu/ Click on “online library” and choose the button called “Education Research”. There are multiple volumes of data there describing numbers of ELLs and ELLs with disabilities, types of programming available, etc. It’s a great resource!
It is possible that universal design will also be included in the next reauthorization of the Individuals with Disabilities Education Act.

The bill for the federal reauthorization of the Individuals with Disabilities Education Act (IDEA) also includes universal design –

**UNIVERSAL DESIGN-** The State educational agency (or, in the case of a district-wide assessment, the local educational agency) shall, to the extent possible, use universal design principles in developing and administering any assessments....

Sec 612, Senate Bill 1248
So, when we think about all students, we have to think about students who have a variety of cultural, physical, sensory, and processing characteristics.

• There are students with all of these characteristics taking both English language proficiency tests and state assessments for accountability.

• Stop here and ask audience: “Who do you see as students who might benefit from universally designed assessments?”

• Perhaps have people think, pair, share here. (Think silently, share their ideas with someone next to them, then share with their table or with the entire group.)

(Follow up with group discussion - Urban/rural, SES, non-literate, ELL with needs but not proper assessment instruments to assess them, vision issues/visual processing – full population of students)
• Before we go any further, it is important to make this critical point. No matter what is done to assessments, if students have not had opportunities to learn the content tested, they will perform poorly.

• Universal design does not water down the standards being tested. ALL students still need to be exposed to these standards and receive quality instruction in order to be able to demonstrate a high level of skill on the assessment. This includes English language learners taking language assessments.
Elements of Universally Designed Assessments

- Inclusive assessment population
- Measures what it intends to measure
- Respects the diversity of the assessment population
- Clear format for text

• Staff at the National Center on Educational Outcomes came up with 7 elements of universally designed assessments. We will describe each of these elements, with examples.

• These elements apply to English language proficiency tests and to content area assessments for accountability.

• And, once again, this is a work in progress. We are continually refining and adding research to this list.

• Some of these elements are currently in place on some assessments. But not all tests take all of these areas into account. Universal design is a way of bringing these important issues to the table for discussion on every assessment.
Elements of Universally Designed Assessments

- Clear pictures and graphics
- Concise and readable text
- Amenable to accommodations

Read through elements – We will be providing description and examples of each element.
Inclusive Assessment Population

✓ Universally designed assessments:

- Consider all types of students in the general curriculum from the beginning

- Include English language learners and students with disabilities in item tryouts and field testing

• First and most important, all students need access to the content being tested. This applies especially to the content area assessments.

• English language learners from the major language groups in a state need to be included in item tryouts and field testing for both English language proficiency tests and general content area assessments.

• Students with a variety of disabilities also need to be included in item tryouts and field testing for both English language proficiency tests and general content area assessments.

• This is difficult because there are very small numbers of students in some of these groups.
Example of Field Test Specifications

• The statewide field test will include students with disabilities and English Language Learners using the accommodations they would normally use in the large-scale assessment.
• There may be need for over sampling of these populations to ensure a valid field test.

Here is an example of a state’s field test specifications that includes students with disabilities using accommodations. Over-sampling small groups is also considered in these specifications.
Measures What it Intends to Measure

✓ *Universally designed assessments reflect good measurement qualities:*

- Reflects the intended content standard (reviewers have information about what is being measured)
- Minimize skills required beyond those being measured

• Language proficiency exams are aligned to ESL standards and content area assessments are aligned to grade level state academic content standards.

• Information about the content standard(s) assessed by each item needs to be supplied to reviewers.

• When considering what is being measured there is somewhat of a “balancing act.” In certain types of test items an additional skill may be necessary. For example, responses to a listening test must be spoken or written, requiring skills in at least one modality in addition to listening. A similar issue is presented on math tests that require skills in reading.

• Careful consideration of the way ESL or content standards are phrased is important in determining which skills involved in responding to an item are extraneous and which are relevant to what is being tested. Try to minimize the effect of the response modality on the modality being assessed.

• If a test has a time limit, careful consideration should be given to why a time limit is necessary for the construct being tested.
Suppose a test item requires a student to read an Aesop’s fable

What could this item measure?

– Decode text?
– Comprehend extended passages?
– Extent of vocabulary?
– Understand the moral or point of the fable?
– Discuss the common elements of any fable?
– Compare and contrast fables with news reports?
– Articulate the relationship between the fable and the overall culture?
– Anything else? (National Center on Accessing the Curriculum, 2003)

If, for example, the purpose of the fable item were to test understanding of the elements commonly found in fables, then supporting word decoding, vocabulary, and comprehension of the story itself would not interfere with the learning challenge. Supports such as text-to-speech, linked vocabulary, or animations illustrating interactions between characters would support different students but still leave the appropriate kind of challenge for all learners. But if the goal were to test decoding, providing those same supports could invalidate the assessment. The reading support would eliminate the students’ opportunity to show reading independence.

This example is from the National Center on Accessing the Curriculum
http://www.cast.org/ncac
Here are some suggestions from researchers.

These are two important points:

1. The description of what each item is testing should be given to item reviewers so that they can assess if design elements of the item might affect the performance of a subgroup of students on that item. Also, it allows them to decide if the construct intended to be tested is actually the one being tested. (e.g. Is this item actually testing math skills for ELLs or their language skills?)

2. What other skills are needed for a student to answer this item besides those it intends to test? Getting at this includes going to students and having them tell you what they are thinking when they are taking the test (think aloud protocol).
Respects the Diversity of the Assessment Population

- Accessible to test takers (consider age, gender, ethnicity, disability, socio-economic level)
- Avoid content that might unfairly advantage or disadvantage any student subgroup

- Avoid regional bias toward or against any group of students that may cause them to have difficulty responding to items or create emotional stress.
- Carefully evaluate what assumptions items make about shared experiences.
- Tests should strive to avoid content that negatively depicts any student subgroup and avoid content that potentially provokes a negative reaction in any student subgroup.
- Age, gender, etc. should not be a barrier to understanding the task an item requires.
According to the National Research Council (1999), bias arises when:

“Deficiencies in the test itself result in different meanings for scores earned by members of different identifiable subgroups.”

It is important to make sure the test is testing the same construct for everyone.
This slide is based on some stories about a commercial testing experience that happened in the southwest, but it brings up an important point for state tests too. You need to carefully consider the population of students in your test in order to discover which items may be less accessible to some groups.

For example, in some Native American tribes such as the Navajo, reptiles are considered a taboo object that students should not look at or be near. In a state or district with high numbers of Navajo students, the test item above could put those students at a disadvantage. The students have to look at the picture of the reptile to answer the question and doing so might upset them or cause them to lose a lot of time. In such a setting with high numbers of Native American students, choosing a different context to measure the same construct would allow the most accessibility to the item. In a state that does NOT have large numbers of Native American students this item might work just fine.

Knowing the students in your area and having members of the largest language communities as part of your item review team will be important in creating accessible test items.
To raise money for a trip to the Wolfridge Environmental Learning Center, sixth graders at Johnson Middle School are selling raffle tickets. The raffle prize is an electric scooter worth more than $300. A total of 500 tickets were sold. You bought two raffle tickets, your sister bought three and your father bought one. What is the probability that someone in your family will win the prize?

• In this math question, the concept of a raffle is not explained. It’s assumed that students will know what one is and how it relates to a prize. If you were missing that knowledge because you had moved here from another country or had religious requirements which prohibited you from participating in activities like this, you might not be able to answer this question.

• If your math standard addressed probability and this item was intended to assess that standard there would be a way to adapt this idea to be less reliant on the knowledge of a “raffle” but still keep the same basic mathematics concept.

• What subgroups of students might have difficulty with this item?
Clear Format for Text

✓ Universally designed assessments use text that can be read quickly and effortlessly

- Standard typeface
- Type size at least 12 point
- Wide spacing
- High contrast
- Sufficient leading (Spacing) between lines of text
- Staggered right margins (no right justification)
- No background graphics behind text

The next section is about the physical appearance of text.

Several examples of effective versus ineffective legibility follow. These examples are from Lighthouse International. http://www.lighthouse.org/print_leg.htm

And from the Accessible Test Department at the American Printing House for the Blind
Typeface:
Standard typeface, with upper and lower case letters, is better than italic, small caps, or all caps

Adapted from: Making Text Legible by A. Arditi
Type Size

All type should be at least 12 point (including captions, keys, labels, and footnotes). Remember that point size differs among typefaces.

Adapted from: Making Text Legible by A. Arditi
Spacing:
Space between letters and between words is wide

This letter spacing is effective.
This letter spacing is not as effective.

This letter spacing is effective.
This letter spacing is not as effective.

Adapted from: Making Text Legible by A. Arditi
Contrast

Text should be printed with the highest possible contrast.

Effective

Not as effective

Adapted from: Making Text Legible by A. Arditi
Leading:
Spacing Between Lines of Text

Leading, or spacing between lines of text, should be at least 25 to 30 percent of the point size. This is because many people with partial sight have difficulty finding the beginning of the next line while reading.

Effective leading

Not effective leading

Adapted from: Making Text Legible by A. Arditi
Staggered Right Margins

Text is easier to read when it is justified on the left and unjustified on the right. This prevents pockets of spaces.

(APH, 2002) Used with permission

This is a picture of a starfish inside a vortex. Of course to you it might look like a giant flower or an octopus in a whirlpool. You might even think it is part or a computer rendering of a bad dream. However this photo is to demonstrate to you how difficult it is to read text when it is placed over a busy background.

(APH, 2002) Used with Permission
These are elements you want to strive for in designing accessible assessments.

- Clear, non-fuzzy pictures
- Dark lines (minimum use of gray scale and shading)
- Sufficient contrast between colors
- Color is not relied on to convey important information or distinctions
- Label pictures and graphics when possible
• When working with reading tests for ELLs or for all students, weigh whether the use of an illustration helps students or interferes with the content being tested. This is a judgment call, but should be carefully considered.
Is this border distracting? Is it necessary?

Once upon a time, there was a little girl who lived with her mother in a small house on the edge of a forest. The girl was called “Little Red Riding Hood” because she wore a coat with a red hood. One day, the girl’s mother called to her and said “Little Red Riding Hood” please take this basket of food to your grandmother. She has been sick and she needs something good to eat.” So Little Red Riding Hood started off through the forest to her grandmother’s house, carrying the basket of food. On the way, she met a wolf. The wolf stepped out that basket of food? “I’m going to my grandmother’s house,” said Little Red Riding Hood. “It’s just on the other side of the forest. My grandmother has been sick and she needs some good food.”
In an actual item, students had to recognize a picture of a fence and use it to answer a question. The picture in the item looked like the picture on the top right of this screen. A Southeast Asian student who was given the item in a “think-aloud” setting hesitated a long time. When asked to describe what she was looking at, she said it was a bridge, not a fence. She was seeing the perspective from the top down, not from the side. She was also relating the picture to what might have been more common in her experience—a wooden footbridge with planks. She looked several times from the text about the fence to the picture of what she thought was a bridge with a worried expression on her face and then finally gave up and said “I can’t do it.”

Context and labels may help identify pictures.
• Here is another thing bias review groups seldom see – instructions at the bottom of the page.

• Do you think the swimmers might affect a student’s performance on a test?
Concise and Readable Text

- Commonly used words
- Vocabulary appropriate for grade level
- Minimum use of unnecessary words
- Idioms avoided unless idiomatic speech is being measured
- Technical terms and abbreviations avoided (or defined) if not related to the content being measured
- Sentence complexity is appropriate for grade level
- Question to be answered is clearly identifiable
• Sometimes bias or sensitivity review teams do not get to see the instructions at the top of the test page – they just review individual items.

• This is the ultimate goal of universally designed assessments along with ensuring that the widest variety of test-takers can respond in the manner intended.
Assessments designed to better include English language learners benefit all types of students!

- Students have the experience to understand the items
- Language is clear, simple and indicates precisely what is required from student (“Plain language”)
- Questions are amenable to supports that ELLs might use
- Cognitive demands are reasonable

Plain language or plain English is a concept that is applied in many fields. If you do an Internet search and type in “plain language” you will find articles from the legal field, from technical editing, from business, etc. They all discuss ways to make your language clear to readers/users. Exactly how that is done can vary depending on the field. There are some education researchers who have studied the use of plain language on assessments and you can find a handout of some suggested reading on plain language in your packet.

Some studies have found that when a state test goes through a plain English review, that plain English test is as beneficial to native English speakers as it is to non-native English speakers.

The following slides describe each of these ideas in more detail.
Abedi research suggests that linguistic complexity of test items is a significant source of measurement error for English language learners (and students with disabilities)

Examples of Linguistic Modifications:

*Familiarity or frequency of non-math vocabulary (unfamiliar or infrequent words changed):*

| Tetris          | Video game |

*Length of nominals (long nominals shortened):*

| Last year's class vice president | The vice president |

(Abedi, et al., 2001)

• Abedi reported on a study he did with simplification of language in NAEP (National Assessment of Educational Progress) items. Here are some examples he gives of ways he simplified language in math items. This can be applicable to all sorts of students—not just ELLs.

• In the first example using the name of a specific computer game was harder for ELLs to understand than when the item was changed to generally talk about video games.

• The sheer amount of language that students have to read through and the complexity of long phrases, clauses and sentences can add to the difficulty of test items even when the students know the content being assessed.

Linguistic Modifications (continued):

*Question phrases (complex question phrases changed to simple question words):*

At which of the following times → When

*Voice of the verb phrase (passive verb forms changed to active):*

If a marble is taken from the bag → If you take a marble from the bag

*Relative clauses (relative clauses removed or recast):*

A report that contains 64 sheets of paper → He needs 64 sheets of paper for each report

• This slide shows other aspects of language simplification (plain language) that can be addressed in test items.

• It is important to carefully consider the type of test and whether language simplification is appropriate for a particular test. Abedi made these recommendations for non-English language arts tests.
In other work, Abedi looked at the effectiveness of a translated math test for ELLs versus a math test with simplified English and came to the conclusion that the simplified English test produced better results for more ELLs than did the translation. So if there is a test that does not assess comprehension of English, linguistic simplification is an option to consider for greater accessibility.

Linguistic Modifications (continued):

Conditional clauses (conditionals either replaced with separate sentences or order of conditional and main clause changed):

If Lee delivers x newspapers

Lee delivers x newspapers

Abedi found that English language learners who received a linguistically modified version of the math test performed significantly better than those receiving the original test items. They also performed significantly faster.
Over 80% of the students also told Abedi they preferred the modified math test because:

- It’s easier to read, and it gets to the point, so you won’t have to waste time
- It’s shorter and doesn’t have, like, complicated words
- I might have a faster time completing that one ‘cuz there’s less reading’

(Abedi, et al., 2001)
Sample Readability Guidelines

- Use simple, clear, commonly used words, eliminating any unnecessary words (Abedi, 2002)
- Clearly define any technical terms that are used
- Break compound complex sentences into several short sentences. State the most important ideas first (Rakow & Gee, 1987)
- Introduce one idea, fact, or process at a time (Rakow & Gee, 1987)

Here, the recommendations of multiple researchers who study issues of language simplification have been combined.
Rebecca Kopriva has produced a document called “Ensuring Accuracy in Testing for English Language Learners.” It’s available on-line through the Council for Chief State School Officers (CCSSO) and is a great resource. (Reference listed in packet of materials.)

These are all recommendations that will aid not only ELLs, but all students in accessing a text.

These should be taken into consideration when designing test items. For example, on an English language proficiency test, you may choose to have a linguistically complex reading passage in order to see if a student can understand it, HOWEVER, the test questions and answer choices (if there are any) should be very simple and clear.

Sentence structure = SVO (subject, verb, object)

Paragraph structure = topic sentence, 2-3 sentences description, conclusion.
Recommendations to Improve Accessibility of Text

- Minimal use of pronouns. Follow a pronoun with the term it refers to in parentheses
- High frequency words
- Avoid words with double meanings and colloquialisms. If used, define them in the text.

From Kopriva (2000)

High frequency words = in an item about pizza, use the word “pepperoni” instead of “anchovies”
This example illustrates some of the difficulties that can occur in translating an item. The words in red were the focus of a comprehension question that followed. In this NCEO study, ELLs who were native Spanish speakers were given the reading passages in English and the test questions and answer choices side by side in English and Spanish. As you look at the next slide, pay attention to how the English comprehension question asks students to make a distinction between some closely related distracters.

In this study, three groups took the test. 2 groups had only questions in English: (1) native English speakers and (2) Spanish-speaking ELLs. The third group consisted of Spanish speaking ELLs that received the bilingual test questions.
Providing a test passage in English, but the questions in a translated form is fairly common among foreign language testing in the US, but not often used in ESL tests. (Anderson, Liu, Swierzbin, Thurlow, & Bielinski, 2000).

For the students who didn’t speak Spanish and didn’t get the Spanish translation, item A was a big distracter.

For the students with the Spanish translation, item A was not a big distracter because the circumlocution used in the translation changed the difficulty of the distracter.

This item or distracter is one that would not easily translate into several some languages.

Researchers advocate creating parallel forms of a test in the common languages needed instead of using translation. (Solano-Flores & Trumbull, 2003).

Translation can be a thorny issue, but one that needs to be addressed up front and the pros and cons weighed.
 Ordering Pizza (Original Item)

The cafeteria manager surveyed the students in a middle school to find out if they would buy Brand X pizza on Friday if the manager sold it. She made a circle graph to display the results of her survey.

Based on the results of the survey, answer the following questions:

1. What fraction of students would buy Brand X pizza on Friday?
2. What percent of students would buy Brand X pizza on Friday?
3. There are 1200 students in this school. How many students will buy Brand X pizza on Friday if the manager's survey is accurate?

This slide and the one that follows show an example of how an item that does NOT assess English language skills can be changed to be more clear. Notice that what the item is assessing stays the same, it’s the setting of the problem and some of the related language that have been simplified.

Ordering Pizza (Revised Item)

Maria surveyed the students in her school to find out if they would like pizza on Friday. She made a circle graph to display the results of her survey.

1. What fraction of students said “yes”?
2. What percent of students said “yes”?
3. There are 1200 students in Maria’s school. How many students said “yes”?

(From Hanson, et al., 1998)

Cognitive Demands

- Amount of text not relevant to items
- Length of text
- Number of long texts
- Timing (may be unspoken)
- Number of unfamiliar words
- Placement of definitions (in text, to side, separate) – if allowed
- Location of native language text – if provided

All of these things add to the cognitive demands of a test. It’s important to look at the whole test or subtest—everything that a student sees.

For example, one long reading passage in a set of fairly short ones might be doable for 4th graders, but 4 reading passages of two pages each might be overwhelming.

The number of unfamiliar words in one passage might seem appropriate but when all of the test passages are put side by side and they each have an equal amount of unfamiliar words, the cognitive demands are heavier for ELLs in particular.

Thought needs to be given to whether cognitive demands of the test are reasonable for those students who have disabilities as well. For example, if a student takes a test in large print, a two page reading passage might become a six page reading passage and make different demands on a vision impaired student than on other students.

Placement of additional information (definitions, native language version of text) on the same page or a facing page can make additional demands on readers who use that information. For example, on a content test that has a side by side bilingual test in English and a student’s native language, the side by side placement of the two versions can make different cognitive demands on a student than having two separate test booklets in two different languages.

Again, as we’ve said before, these are issues to be discussed and debated, not hard and fast rules about things to avoid.
Amenable to Accommodations

- Universally designed assessments allow needed accommodations to be used

- Plan for students who continue to need accommodations

- Facilitate the use of accommodations such as bilingual dictionaries or translations (when appropriate), assistive technology, oral presentation, large print, Braille

• Some students will continue to need accommodations.
• It is important to plan for the use of these accommodations.
• Involve members of the major language groups on item review committees. Ask them for input on whether particular items will translate into those languages. Consideration of these issues prior to administration of a test will also help with the administration of oral interpretations in the native language, if allowed on a content test. This issue is also relevant for sign language interpretations of tests, when appropriate.
• There are many ways to translate content area assessments, such as side-by-side or developing parallel forms. Carefully consider the plusses and minuses of each way prior to making a decision about your state test. There is no perfect solution.
• Universally designed assessments make it easier for more students to use the same test format because many accommodations (e.g. large-print, translations) can be built right into the design.
• Students who could benefit from the use of an accommodation (e.g. translation, glossary) sometimes do not use the accommodation in front of their peers. (Anderson, Liu, Swierzbin, Thurlow, Bielinski, 2000).
• For an example of a computer-based web page that is amenable to a screen reader vs. one that is not, go to: http://www.humanfactors.com/downloads/chocolateaudio.asp
15. The hike along the Appalachian Trail from Kinsman Pond (top left of map) to Lonesome Lake (near the top of the map) can be best described as
A. an ascent, then a descent.
B. an ascent, then level.
C. a descent, then level.
D. a descent, then an ascent.
Here is a revised graphic that would clarify the item.

MORE ACCESSIBLE VERSION

Franconia Notch Trails

Key
- Appalachian Trail
- pond or lake
- contour line

contour interval = 200 ft.

15. The hike along the Appalachian Trail from Kinsman Pond (top left of map) to Lonesome Lake (near the top of the map) can be best described as
   A. an ascent, then a descent.
   B. an ascent, then level.
   C. a descent, then level.
   D. a descent, then an ascent.

Items and revision used by permission from the state of NH and DRC

This graphic clarifies the graphic and makes it more amenable to accommodations by getting rid of information that is not crucial to answering the question.
Universally designed assessments consider the design of the response venue as well as the assessment itself

- Large bubbles that avoid most challenges of low vision or difficulty with fine motor skills (Grise, Beattie, Algozzine, 1982)
- Consideration of age of students in selecting format (avoid separate answer sheets for younger students) (Tindal and others suggest 4th grade as transition point).

Reviewers should also have a chance to look at response formats that will be presented on the actual test.
Clear Response Grid

Hard to see

Parts of a Response Grid
For Grade 8, response grids have the following parts:

- answer boxes
- fraction bar
decimal point

Larger bubbles.
Easier to see.

Gray columns changed to yellow.

Parts of a Response Grid

- answer boxes
- fraction bars
decimal points

number bubbles

(APH, 2003) Used with permission

Be sure to also review the clarity of the response grid.
http://www.aph.org
Computer-Based Testing

All of the Elements of Universal Design
Plus Some Additional Opportunities and Challenges

Everything we have talked about so far applies to computer based tests
Students may be more likely to use the accommodations that may help them when they are built into a test on a computer that can be individualized (e.g. translations, glossaries) (Anderson, Liu, Swierzbin, Thurlow, & Bielinski, 2000).

Opportunities

- Efficient administration
- Preferred by students
- Improved writing performance
- Built-in accommodations
- Immediate results
- Efficient item development
- Increased authenticity
Challenges

- Use of technology cannot take the place of content mastery
- Issues of equity and skill in computer use
- Added challenges for some students
- Technological challenges
- Security of online data
- Lack of expertise in designing accessible Web pages
- Prohibitive development cost

Each of these challenges need to be carefully considered in the development of computer based tests
Example: Kentucky Online Assessment

• Web-based, individualized assessment:
  – Students with IEP or 504 Plan that specifies need for "reader" as an instructional and assessment accommodation;
  – Students who require and routinely use text-reader or screen-reader technologies to access printed material in classroom instruction and assessment;
  – Students who have accessed and used the CATS Online Practice Area.

• Based upon success of pilot studies, 16 districts, 31 Schools, & 204 students participated in “live” CATS Online in the spring of 2003.

• Kentucky administered its state test online for students with disabilities for the first time in spring 2003.

• Show example of screen reader for web sites with different designs.

• For an example of a computer-based web page that is amenable to a screen reader vs. one that is not, go to: http://www.humanfactors.com/downloads/chocolateaudio.asp (If did not show earlier with amenable accommodations slide).
CATS Online

Here’s an example of a screen from Kentucky’s test.

http://www.education.ky.gov/KDE/Administrative+Resources/Testin
g+and+Reporting+/District+Support/CATS+Online+Assessment/def
ault.htm
How to Develop Universally Designed Assessments

- Develop frameworks for alignment of items with content
- Include elements of universally designed assessments in test specifications
- Develop guidelines for item developers to follow

Each of these suggestions will help ensure of universally designed assessments
How to Develop Universally Designed Assessments

Flag items with bias and design issues:

- Consult with end users – students representing all subgroups
- Consult with bias or sensitivity review teams – representing all subgroups
- Analyze field test data and statewide test results by item and subgroup

There are 3 ways to flag items for bias and design issues
Here are some of the things test and item reviewers need in order to perform a careful and comprehensive review.

Items often go through a review before they are field tested in order to improve the items. At this point one would not have field test results at hand. However, after they are field tested and items are reviewed again, item reviewers should have the results of the test.
Here are some of the people who make up an effective review team. Can you think of any one else who might be involved?

**Review Team Members**

- Grade-level content experts
- Representatives of major cultural groups
- Representatives of major disability groups
- Second language acquisition experts
- **Who else?**
Remember!

• This is a work in progress

• Considerations put important issues on the table

• Ongoing input is important!
These are the materials that are included in the packet that accompanies this training module.
Visit:  education.umn.edu/nceo

or Search for NCEO

Click on “Universal Design”
Considerations for Universally Designed Assessment Items

These guidelines contain suggestions that test developers, item reviewers, and others working on the development of assessments should consider at the beginning stages of designing an assessment that is accessible to the widest range of students possible. Tests should be reviewed for cultural appropriateness by community specialists representing the major student subgroups being tested, as well as by ESL and academic content specialists.

<table>
<thead>
<tr>
<th>Consider how the item:</th>
<th>English Language Proficiency Listening/ Speaking Computer Based</th>
<th>English Language Proficiency Reading and Writing Paper/Pencil</th>
<th>Grade level Academic Content Area Assessment Computer Based</th>
<th>Grade level Academic Content Area Assessment Paper/Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures what it intends to measure</td>
<td>• Reflects the intended content standard (reviewers have information about the content being measured)</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
</tr>
<tr>
<td></td>
<td>• Minimize skills required beyond those being measured</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
</tr>
</tbody>
</table>

Notes:

a. Language proficiency exams are aligned to ESL standards and content area assessments are aligned to grade level state academic content standards.

b. Information about the content standard(s) assessed by each item needs to be supplied to reviewers.

c. When considering what is being measured there is somewhat of a “balancing act.” In certain types of test items an additional skill may be necessary. For example, responses to a listening test must be spoken or written, requiring skills in at least one modality in addition to listening. A similar issue is presented on math tests that require skills in reading.

d. Careful consideration of the way ESL or content standards are phrased is important in determining which skills involved in responding to an item are extraneous and which are relevant to what is being tested. Try to minimize the effect of the response modality on the modality being assessed.

e. If a test has a time limit, careful consideration should be given to why a time limit is necessary for the construct being tested.
Consider how the item:

- **English Language Proficiency Listening/Speaking**
- **Computer Based**
- **English Language Proficiency Reading and Writing**
- **Paper/Pencil**
- **Grade level Academic Content Area Assessment**
- **Computer Based**
- **Grade level Academic Content Area Assessment**
- **Paper/Pencil**

### Respects the diversity of the assessment population

- Accessible to test takers (consider age, gender, ethnicity, and socio-economic level)
- Avoids content that might unfairly advantage or disadvantage any student subgroup

<table>
<thead>
<tr>
<th></th>
<th>English Language Proficiency Listening/Speaking Computer Based</th>
<th>English Language Proficiency Reading and Writing Paper/Pencil</th>
<th>Grade level Academic Content Area Assessment Computer Based</th>
<th>Grade level Academic Content Area Assessment Paper/Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>♦</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
</tr>
</tbody>
</table>

### Notes:

- a. Avoid regional bias toward or against any group of students that may cause them to have difficulty responding to items or create emotional stress.
- b. Carefully evaluate what assumptions items make about shared experiences.
- c. Tests should strive to avoid content that negatively depicts any student subgroup and avoid content that potentially provokes a negative reaction in any student subgroup.
- d. Gender, etc. should not be a barrier to understanding the task an item requires.
Consider how the item:

<table>
<thead>
<tr>
<th>Item</th>
<th>English Language Proficiency</th>
<th>English Language Proficiency</th>
<th>Grade level Academic Content Area Assessment</th>
<th>Grade level Academic Content Area Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Standard typeface</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
<tr>
<td>• Type size appropriate for age group (12 point minimum for all print, including captions, footnotes, and graphs)</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
<tr>
<td>• Wide spacing between letters, words, and lines</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
<tr>
<td>• High contrast between color of text and background</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
<tr>
<td>• Sufficient leading (blank space) between lines of text</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
<tr>
<td>• Staggered right margins (no right justification)</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
</tbody>
</table>

Notes:

a. Use common fonts such as Times or Arial.
b. Avoid decorations and flourishes.
c. For computer-based tests, use relative font sizing instead of a fixed font size because of variations in monitors, resolutions, and settings.
d. The term “blank space” rather than “white space” may be more accurate because the background is not always white.
**Consider how the item:**

<table>
<thead>
<tr>
<th>English Language Proficiency Listening/Speaking Computer Based</th>
<th>English Language Proficiency Reading and Writing Paper/Pencil</th>
<th>Grade level Academic Content Area Assessment Computer Based</th>
<th>Grade level Academic Content Area Assessment Paper/Pencil</th>
</tr>
</thead>
</table>

**Has clear pictures and graphics (when essential to item)**

- Pictures are needed to respond to item
- Clear, non-fuzzy pictures
- Dark lines (minimum use of gray scale and shading)
- Sufficient contrast between colors
- Color is not relied on to convey important information or distinctions
- Label pictures and graphics when possible.

**Notes:**

a. When working with reading tests for ELLs or for all students, weigh whether the use of an illustration helps students or interferes with the content being tested. This is a judgment call, but should be carefully considered.

b. Labeling pictures, when possible, is helpful. This is true even if the picture seems obvious. For example, if an item asks students to look at a picture of a fence from a side view, a student from another background may look at the picture top down and see a foot bridge.

c. Carefully consider the importance of color distinctions for answering an item. At present, this issue is perhaps more important in the construction of computer-based tests that may use color to a much greater extent than paper/pencil tests.

d. Give additional clues besides color when possible (e.g., use text label “stop” and “go” next to stop light images).
### Consider how the item:

<table>
<thead>
<tr>
<th>Has concise and readable text</th>
<th>English Language Proficiency Listening/ Speaking Computer Based</th>
<th>English Language Proficiency Reading and Writing Paper/Pencil</th>
<th>Grade level Academic Content Area Assessment Computer Based</th>
<th>Grade level Academic Content Area Assessment Paper/Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Commonly used words</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
<tr>
<td>• Vocabulary appropriate for grade level</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
<tr>
<td>• Minimum use of unnecessary words</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
<tr>
<td>• Idioms avoided unless idiomatic speech is being measured</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
<tr>
<td>• Technical terms and abbreviations avoided (or defined) if not related to the content being measured</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
<tr>
<td>• Sentence complexity is appropriate for grade level</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
<tr>
<td>• Question to be answered is clearly identifiable</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
<td>✦</td>
</tr>
</tbody>
</table>

### Notes:

a. The use of common words depends on the content assessed. For example, if vocabulary is being tested, difficult or uncommon words might be appropriate to include.
b. Sentence structure and vocabulary should not present an unnecessary challenge to comprehension.
c. English language learners may know less common words but may not know phrasal words. It is difficult to assume what is uncommon or difficult.
d. If reading is not the primary construct tested, keep reading level at or below grade level.
e. Readability indices are now found in major word processors, however, it is important to check to see they are working as intended.
f. Determination of complexity can include many factors such as: use of clauses, use of the passive, number of syllables in a word, length of sentences, length of single passage, combined length of all reading passages, amount of extraneous text involved in non-reading problems, etc. Complexity needs to be considered on all tests.
g. When using authentic texts, complexity may be difficult to control on reading passages, but should be considered on test questions.
Consider how the item:

<table>
<thead>
<tr>
<th>Allows changes to its format without changing its meaning or difficulty (including visual or memory load)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Allows for the use of Braille or other tactile format</td>
</tr>
<tr>
<td>• Allows for signing to a student</td>
</tr>
<tr>
<td>• Allows for the use of oral presentation to a student</td>
</tr>
<tr>
<td>• Allows for the use of assistive technology</td>
</tr>
<tr>
<td>• Allows for translation into another language</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>English Language Proficiency Listening/ Speaking Computer Based</th>
<th>English Language Proficiency Reading and Writing Paper/Pencil</th>
<th>Grade level Academic Content Area Assessment Computer Based</th>
<th>Grade level Academic Content Area Assessment Paper/Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>♦</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
<td></td>
</tr>
</tbody>
</table>

Notes:

a. Check to see which tests allow oral presentation.
b. Consider translation of directions on a language proficiency test.
c. Involve members of the major language groups in item review committees.
d. Consideration of these issues prior to administration of a test will also help with the administration of oral interpretations in the native language, if allowed on a content test. This issue is also relevant for sign language interpretations of tests, when appropriate.
e. There are many ways to translate content area assessments, such as side-by-side or developing parallel forms. Carefully consider the plusses and minuses of each way prior to making a decision about your state test. There is no perfect solution.
Overall test considerations
Consider how the test:

<table>
<thead>
<tr>
<th>Has an overall appearance that is clean and organized</th>
</tr>
</thead>
<tbody>
<tr>
<td>• All images, pictures and text provide information necessary to respond to the item</td>
</tr>
<tr>
<td>• Information is organized in a manner that is consistent with an academic English framework with a left-right, top-bottom flow.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>English Language Proficiency Listening/Speaking Computer Based</th>
<th>English Language Proficiency Reading and Writing Paper/Pencil</th>
<th>Grade level Academic Content Area Assessment Computer Based</th>
<th>Grade level Academic Content Area Assessment Paper/Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
</tr>
</tbody>
</table>

Notes:

a. Images, pictures and text that may not be necessary include sidebars, overlays, callout boxes, visual crowding, shading, and general busyness - anything that may distract a student.
b. Pictures should have a purpose other than simply to be decorative.
c. Carefully consider whether students from some groups may misinterpret the flow of text and graphics based on characteristics of their native language or culture. Left-right and top-bottom flow is cultural. In some languages, text may flow top to bottom, for example.
d. When using “authentic visuals” (e.g., a map), carefully consider what is being tested in addition to the intended content.
Consider how:

<table>
<thead>
<tr>
<th>Considerations for computer-based assessments</th>
<th>English Language Proficiency Listening/Speaking &lt;br&gt;Computer Based</th>
<th>English Language Proficiency Reading and Writing &lt;br&gt;Paper/Pencil</th>
<th>Grade level Academic Content Area Assessment &lt;br&gt;Computer Based</th>
<th>Grade level Academic Content Area Assessment &lt;br&gt;Paper/Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Layout and design</strong></td>
<td>• Sufficient contrast between background and text and graphics for easy readability.</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
</tr>
<tr>
<td></td>
<td>• Color is not relied on to convey important information or distinctions.</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
</tr>
<tr>
<td></td>
<td>• Font size and color scheme can be easily modified (through browser settings, style sheets or on-screen options).</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
</tr>
<tr>
<td></td>
<td>• Stimulus and response options are viewable on one screen when possible.</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
</tr>
<tr>
<td></td>
<td>• Page layout is consistent throughout the test.</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
</tr>
<tr>
<td><strong>Navigation</strong></td>
<td>• Navigation is clear and intuitive. It makes sense and is easy to figure out.</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
</tr>
<tr>
<td></td>
<td>• Navigation and response selection is possible by mouse click or keyboard.</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
</tr>
<tr>
<td></td>
<td>• Option to return to items and return to place in test after breaks</td>
<td>♦</td>
<td>♦</td>
<td>♦</td>
</tr>
</tbody>
</table>
### Screen reader considerations
- Item is intelligible when read by a text/screen reader.
- Links make sense when read out of visual context. ("Go to the next question" rather than ‘click here”).
- Non-text elements have a text equivalent or description.
- Tables are only used to contain data, and make sense when read by screen reader.

### Test specific options
- Access to other functions is restricted (e.g. email, internet, instant messaging).
- Pop up translations and definitions of key words/phrases are available if appropriate to the test.

### Computer capabilities
- Adjustable volume.
- Speech recognition available (to convert user’s speech to text.
- Test is compatible with current screen reader software.

### Notes:
**Layout and design**
a. The design of the color of the test needs to take into account those test takers with color blindness including red/green distinctions.
b. The electronic format needs to be accessible through the specific assistive technology the student has experience using during the testing. The latest technology may not be what is used in the schools.

**Navigation**
a. How to navigate and navigation symbols should be intuitive and/or explained at the beginning of a test.
b. The screen resolution varies on different computers. Reviewers should check out items on different types of computers commonly used in schools.
c. Schools need reasonable minimum standards for computer and audio requirements for test.
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>d.</strong> Test administration instructions should include standardized settings for the computer.</td>
<td></td>
</tr>
<tr>
<td><strong>e.</strong> Students need practice opportunities before taking computer-based tests.</td>
<td></td>
</tr>
<tr>
<td><strong>f.</strong> Some listening tests may want to limit the number of times a student can listen to a recording, depending on standards being tested.</td>
<td></td>
</tr>
</tbody>
</table>

**Screen reader**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a.</strong> Images and animations have text labels, IF this does not supply the answer.</td>
<td></td>
</tr>
<tr>
<td><strong>b.</strong> Captioning and transcripts of audio and video are available.</td>
<td></td>
</tr>
<tr>
<td><strong>c.</strong> Provide titles and summaries for tables and graphs</td>
<td></td>
</tr>
<tr>
<td><strong>d.</strong> Header cells for columns and/or rows are designated.</td>
<td></td>
</tr>
<tr>
<td><strong>e.</strong> Information in tables makes sense when linearized (i.e. read top left to bottom right cell).</td>
<td></td>
</tr>
<tr>
<td><strong>f.</strong> Current screen reader technology might be difficult for ELLs to understand, real voice technology may be needed</td>
<td></td>
</tr>
</tbody>
</table>

**Test specific options**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a.</strong> Access to spell check might also be limited depending on the test.</td>
<td></td>
</tr>
<tr>
<td><strong>b.</strong> Variable audio speed might be useful to some students if it does not interfere with the standard being tested.</td>
<td></td>
</tr>
</tbody>
</table>
**Supporting Statements by Researchers**

<table>
<thead>
<tr>
<th>Measures what it intends to measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Test development begins with a careful consideration of the skills proposed for measurement (Popham &amp; Lindheim, 1980).</td>
</tr>
<tr>
<td>• Language in non-language arts assessments needs to be “transparent” enough to students to clearly determine constructed being measured (Sharrocks-Taylor &amp; Haregreaves, 1999).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Respects the diversity of the assessment population</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Items must be reviewed for bias that may exist against particular populations (National Research Council, 1999).</td>
</tr>
<tr>
<td>• Items that are designed from the start with equity and accessibility features are less likely to be biased against particular populations (Kopriva, 2000).</td>
</tr>
<tr>
<td>• Items must be free of content that makes a student’s socioeconomic status or inherited academic aptitudes the dominant influence on how a student’s will respond to the item (Popham, 2001).</td>
</tr>
<tr>
<td>• Items must be free of content that may unfairly benefit or penalize students from diverse ethnic, socioeconomic, or linguistic backgrounds, or students with disabilities (Popham, 2001).</td>
</tr>
<tr>
<td>• Cultural norms, beliefs, and customs need to be respectfully reflected in illustrations (Schiffman, 1995).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Has a clear format for text</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The point sizes most often used are 10 and 12 point for documents to be read by people with excellent vision reading in good light (Gaster &amp; Clark, 1995).</td>
</tr>
<tr>
<td>• Fourteen point type increases readability and can increase test scores for both students with and without disabilities, compared to 12-point type (Fuchs, Fuchs, Eaton, Hamlett, Binkley, &amp; Crouch, 2000).</td>
</tr>
<tr>
<td>• Type size for captions, footnotes, keys, and legends needs to be at least 12 point (Arditi, 1999).</td>
</tr>
<tr>
<td>• Larger type sizes are most effective for young students who are learning to read and for students with visual difficulties (Hoerner, Salend, &amp; Kay, 1997).</td>
</tr>
<tr>
<td>• Large print is beneficial for reducing eye fatigue (Arditi, 1999).</td>
</tr>
<tr>
<td>• Shapes of letters and numbers should enable people to read text “quickly, effortlessly, and with understanding”</td>
</tr>
</tbody>
</table>
• The relationship between readability and point size is also dependent on the typeface used (Gaster & Clark, 1995; Worden, 1991).
• Letters that are too close together are difficult for partially sighted readers. Spacing needs to be wide between both letters and words (Gaster & Clark, 1995).
• Fixed-space fonts seem to be more legible for some readers than proportional-spaced fonts (Gaster & Clark, 1995).
• Leading should be 25-30 percent of the point (font) size for maximum readability (Arditi, 1999).
• Leading alone does not make a difference in readability as much as the interaction between point size, leading and line length (Worden, 1991).
• Standard typeface, upper and lower case, is more readable than italic, slanted, small caps, or all caps (Tinker, 1963).
• Text printed completely in capital letters is less legible than text printed completely in lower-case, or normal mixed-case text (Carter, Dey & Meggs, 1985).
• Italic is far less legible and is read considerably more slowly than regular lower case (Worden, 1991).
• Boldface is more visible than lower case if a change from the norm is needed (Hartley, 1985).
• Staggered right margins are easier to see and scan than uniform or block style right justified margins (Arditi, 1999; Grise et al., 1982; Menlove & Hammond, 1998).
• Justified text is more difficult to read than unjustified text - especially for poor readers (Gregory & Poulton, 1970; Zachrisson, 1965).
• Justified text is also more disruptive for good readers (Muncer, Gorman, Gorman, & Bibel, 1986).
• A flush left/ragged right margin is the most effective format for text memory. (Thompson, 1991).
• Unjustified text may be easier for poorer readers to understand because the uneven eye movements created in justified text can interrupt reading (Gregory & Poulton, 1970; Hartley, 1985; Muncer, Gorman, Gorman, & Bibel, 1986; Schriver, 1997).
• Justified lines require the distances between words to be varied. In very narrow columns, not only are there extra wide spaces between words, but also between letters within the words (Gregory & Poulton, 1970).
• Longer lines, in general, require larger type and more leading (Schriver, 1997).
• Optimal length is 24 picas - about 4 inches (Worden, 1991).
• Lines that are too long make readers weary and may also cause difficulty in locating the beginning of the next line, causing readers to lose their place (Schriver, 1997; Tinker, 1963).
• Lines of text should be about 40-70 characters, or roughly eight to twelve words per line (Heines, 1984; Osborne, 2001; Schriver, 1997).
• Blank space anchors text on the paper and helps increase legibility (Menlove & Hammond, 1998; Smith & McCombs, 1971).
• A general rule is to allow text to occupy only about half of a page. Too many test items per page can make items difficult to read (Tinker, 1963).

Has clear pictures and graphics (when essential to item)

• Graphics with a clear sense of unity, a clear focal point, and balance reduce the cognitive load of perceiving information and. Computer-based tests should allow students to change the size of thus increase the font (see computer specific considerations below). speed with which the user can access graphic material (Szabo and Kanuka, 1998).
• If illustrations are present they are at best essential information, good if they support the information, and unnecessary if they are unrelated to the construct or item (Sharrock-Taylor and Hargreaves, 1999).
• Illustrations should be placed directly next to the information for which they refer (Silver, 1994; West, 1997).
• Placing labels directly on plot lines of graphs reduces the load on short-term memory (Gregory & Poulton, 1970).
• Quantitative displays should be structured so that readers can easily construct appropriate inferences about the data (Schrider, 1997).
• Illustrations should be placed directly next to the information for which they refer (Silver, 1994; West, 1997).
• Graphs, illustrations, and other graphic aids can facilitate comprehension (Rakow & Gee, 1987)

Has concise and readable text

• General readability principles such as fewer words per sentence and the removal of irrelevant difficult words increases comprehension of items (Popham & Lindhein, 1980; Rakow & Gee, 1987).
• Flow of sentences is also an important feature. Caution should be taken when reducing reading load so that sentences do not become disjointed or incomprehensible (Anderson, Hiebert, Scott & Wilkinson, 1985).
• Compound sentences can be written in two separate sentences (if sentences are still comprehensible) (Gaster & Clarke, 1995).
• Most important ideas should be stated first in a sentence (Gaster & Clarke, 1995).
• Noun-pronoun relationships should be clear (Gaster & Clarke, 1995).
• Illustrations should be placed close to the text they support (Gaster & Clarke, 1995), or removed if they do not support text.
• Readability increases when students have likely had experiences or prior knowledge relating to items (Rakow & Gee, 1987).
• Content within items is clearly organized (Rakow & Gee, 1987)
• Questions are clearly framed (Rakow & Gee, 1987)
- Limit the number of words, difficulty of words, and grammatical complexity of test materials (Popham & Lindheim, 1980)
- Simple, clear, commonly used words should be used whenever possible (Gaster & Clarke, 1995).
- Technical terms should be defined (Gaster & Clarke, 1995).
- One idea, fact, or process should be introduced at a time, then ideas developed logically (Gaster & Clarke, 1995).
- If time and setting are important to the sentence, they should be placed at the beginning of the sentence (Gaster & Clarke, 1995).
- Sequence steps of instructions in the exact order that they will be needed (Gaster & Clarke, 1995).
- Vocabulary should be grade-level appropriate (Rakow & Gee, 1987).
- Sentence complexity must be appropriate for grade level (Rakow & Gee, 1987).
- Definitions and examples must be clear and understandable (Rakow & Gee, 1987).
- Required reading skills are appropriate for students’ cognitive level (Rakow & Gee, 1987).
- Use of plain language: “text-based language that is straightforward, concise, and uses everyday words to convey meaning. The goal of plain language editing strategies is to improve the comprehensibility of written text while preserving the essence of its message.” (Hanson, Hayes, Schriver, LeMahieu, & Brown, 1998, p.2)
- Reduce the verbal and organizational complexity of test items while preserving their essential content (i.e., the skills and concepts they were intended to measure)” (Hanson et al, 1998, p.2).
- Reduce Excessive Length: Reduce wordiness and remove irrelevant material (Brown, 1999).
- Eliminate unusual or low frequency words and replace with common words (ex. replace ‘utilize’ with ‘use’). (Brown, 1999)
- Avoid ambiguous words (ex. crane) (Brown, 1999)
- Avoid irregularly spelled words (ex. trough, feign) (Brown, 1999)
- Avoid proper names, replace with common names or no names at all (Brown, 1999)
- Avoid inconsistent naming or graphic conventions (ex. multiple names for the same concept - unless recognizing multiple names for the same concept is the construct being tested) (Brown, 1999)
- Avoid unclear signals about how to direct attention (ex. phrases such as “in the table below” are helpful) (Brown, 1999)
- Mark all questions clearly (Brown, 1999)
- Simplified English on math tests enables comparable performance by English Language Learners and English proficient students (Abedi, Hofstetter, Baker, & Lord, 2001)
- Linguistic demands of tests often pose the greatest barrier to students’ ability to demonstrate knowledge of mathematical concepts (Hanson, 1997).
- Plain language versions of tests assist students that understand content but are less proficient in English language. They do not help students who do not understand content (Brown, 1999).
**Allows changes to its format without changing its meaning or difficulty (including visual or memory load)?**

- Construct irrelevant graphs, vertical text, untranslatable material, and decorative graphics all create situations where accommodating students who use Braille, American Sign Language or non-English languages is difficult.

**Additional considerations for computer-based assessments**

- Students reported difficulties with computers including: excessive need for forward and back buttons, unfamiliarity with response mechanisms, and an inability to see entire problems on screens (Trotter, 2001).
- Students may not be familiar with skills like scrolling or using text on multiple screens (Cole, Tindal & Glasgow, 2000).
- Some students have had little access to computers and calculators prior to testing (Bridgeman, Harvey & Braswell, 1995; MacArthur & Graham, 1987).
References


Heubert, J.P. (2002). *Disability, race, and high-stakes testing of students*. Teachers College, Columbia University; Columbia Law School: National Center for Accessing the General Curriculum.


## Considerations for Universally Designed Assessment Items

<table>
<thead>
<tr>
<th>Consider how the item:</th>
<th>Areas of Strength</th>
<th>Areas of Concern</th>
<th>Suggestions for Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measures what it intends to measure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Reflects the intended content standard (reviewers have information about the content being measured)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Minimize skills required beyond those being measured</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Respects the diversity of the assessment population</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Accessible to test takers (consider age, gender, ethnicity, and socio-economic level)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Avoids content that might unfairly advantage or disadvantage any student subgroup</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Has a clear format for text</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Standard typeface</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Type size appropriate for age group (12 point minimum for all print, including captions, footnotes, and graphs)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Wide spacing between letters, words, and lines</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• High contrast between color of text and background</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Sufficient leading (blank space) between lines of text</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Staggered right margins (no right justification)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Has clear pictures and graphics (when essential to item)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Pictures are needed to respond to item</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Clear, non-fuzzy pictures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Dark lines (minimum use of gray scale and shading)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Sufficient contrast between colors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Color is not relied on to convey important information or distinctions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Label pictures and graphics when possible.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consider how the item:</td>
<td>Areas of Strength</td>
<td>Areas of Concern</td>
<td>Suggestions for Improvement</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td><strong>Has concise and readable text</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Commonly used words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Vocabulary appropriate for grade level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Minimum use of unnecessary words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Idioms avoided unless idiomatic speech is being measured</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Technical terms and abbreviations avoided (or defined) if not related to the content being measured</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Sentence complexity is appropriate for grade level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Question to be answered is clearly identifiable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Allows changes to its format without changing its meaning or difficulty (including visual or memory load)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Allows for the use of Braille or other tactile format</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Allows for signing to a student</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Allows for the use of oral presentation to a student</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Allows for the use of assistive technology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Allows for translation into another language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Has an overall appearance that is clean and organized</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• All images, pictures and text provide information necessary to respond to the item</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Information is organized in a manner that is consistent with an academic English framework with a left-right, top-bottom flow.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Considerations for Universally Designed Assessment Items

<table>
<thead>
<tr>
<th>Star (*) areas of strength and Check (√) areas of concern for each item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measures what it intends to measure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Reflects the intended content standard (reviewers have information about the content being measured)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Minimize skills required beyond those being measured</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Respects the diversity of the assessment population</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Accessible to test takers (consider age, gender, ethnicity, and socio-economic level)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Avoids content that might unfairly advantage or disadvantage any student subgroup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Has a clear format for text</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Standard typeface</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Type size appropriate for age group (12 point minimum for all print, including captions, footnotes, and graphs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Wide spacing between letters, words, and lines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• High contrast between color of text and background</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Sufficient leading (blank space) between lines of text</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Staggered right margins (no right justification)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Has clear pictures and graphics (when essential to item)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Pictures are needed to respond to item</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Clear, non-fuzzy pictures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Dark lines (minimum use of gray scale and shading)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Sufficient contrast between colors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Color is not relied on to convey important information or distinctions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Label pictures and graphics when possible.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consider how the item:</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----</td>
</tr>
<tr>
<td>Has concise and readable text</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Commonly used words</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Vocabulary appropriate for grade level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Minimum use of unnecessary words</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Idioms avoided unless idiomatic speech is being measured</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Technical terms and abbreviations avoided (or defined) if not related to the content being measured</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Sentence complexity is appropriate for grade level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Question to be answered is clearly identifiable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allows changes to its format without changing its meaning or difficulty (including visual or memory load)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Allows for the use of Braille or other tactile format</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Allows for signing to a student</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Allows for the use of oral presentation to a student</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Allows for the use of assistive technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Allows for translation into another language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has an overall appearance that is clean and organized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• All images, pictures and text provide information necessary to respond to the item</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Information is organized in a manner that is consistent with an academic English framework with a left-right, top-bottom flow.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Revised items for use in training

The following two sample test items have gone through a review process and have been revised using the considerations for universal design. The intent of the revisions is to illustrate how the items can be improved to be more accessible to students.

These items are meant to be used as examples in training. You are encouraged to use actual items or released items from your own state test in training so that participants can use the considerations for universal design in a hands-on activity.

We thank Tim Kurtz, New Hampshire Department of Education; Marge Petit, Center for Assessment in New Hampshire; Carl Lager, Vermont State Department of Education; and the staff at Data Recognition Corporation for their collaboration in revising these items as well as several of the items used in the PowerPoint presentation included in this module.
Standard: Student will be able to read bar graphs, line graphs and tables. Student will be able to pose questions for scientific investigations and make predictions about the outcomes.

Original item...

Use the table below to answer question 28.

<table>
<thead>
<tr>
<th>Ball</th>
<th>Diameter (cm)</th>
<th>Mass (g)</th>
<th>Release height (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>50</td>
<td>1</td>
</tr>
</tbody>
</table>

28. To demonstrate how energy is transferred, two balls are dropped into a bucket full of Jell-O. The balls are released at the same time and reach the surface of the Jell-O at the same time. Which result is most likely?

A. Ball A will go deeper into the Jell-O than Ball B.
B. Ball B will go deeper into the Jell-O than Ball A.
C. Both balls will go the same distance into the Jell-O.
D. Both balls will bounce off the surface of the Jell-O.

NH Grade 6 Science – May 2003
More accessible...

Use the table below to answer question 28.

<table>
<thead>
<tr>
<th>Ball</th>
<th>Diameter (cm)</th>
<th>Mass (g)</th>
<th>Release height (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>2</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>Y</td>
<td>2</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

29. To demonstrate how energy is transferred, two balls are dropped into a bucket full of sand. The balls are released at the same time and reach the surface of the sand at the same time. Which result is most likely?

A. Ball X will go deeper into the sand than Ball Y.
B. Ball Y will go deeper into the sand than Ball X.
C. Both balls will go the same distance into the sand.
D. Both balls will not go into the sand.

Changes made:
- Variables changed to X and Y so not to be confused with answer choices.
- Sand used since some ESL students not familiar with Jell-O or what it is.
- Masses changed to be more realistic.
Standard:
- Students will use an assortment of measuring instruments, with a variety of scales, such as rulers, thermometers, graduated cylinders, and timers.
- Students will be able to record data using appropriate units.

Changes made:
- Change “Leaf A” to “Leaf X” to avoid confusion with options A,B,C,D.
- Move ruler and add dotted line to remove obstacle of interpreting the location of the leaf tip.
- Make numbers on the ruler larger and easier to read.
- Remove overlay ruler which is difficult to read.
- Remove pot to make it clear that measurement is from soil line to leaf tip.