

Creating Student-

It's hard to create accessible tests that help students show what they know. Here's some how-to.

Spencer J. Salend

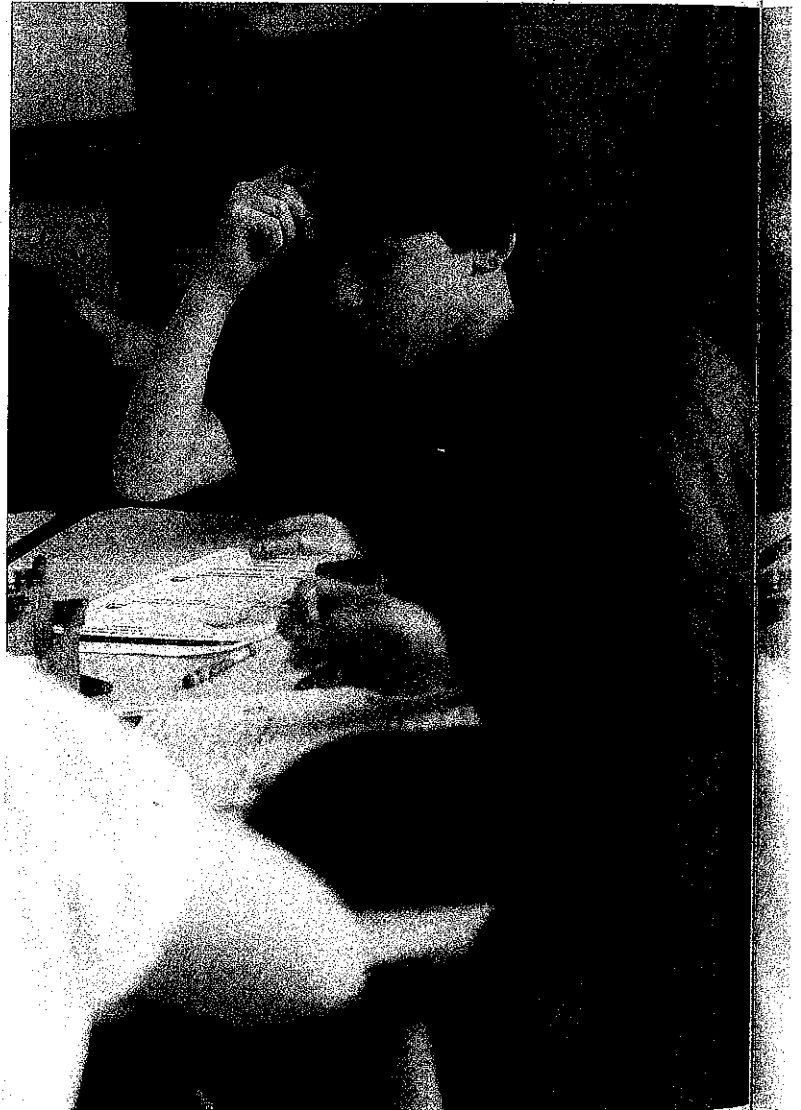
The teaching team at Madison Middle School was concerned about students' test scores. Although the students' classroom performance indicated that they grasped the concepts being taught, their scores on teacher-made tests indicated quite the opposite. Determined to understand what was happening, the teachers asked students to anonymously share what they thought about recent tests and why they had received the grades they did. Students' comments included the following:

- "The tests don't cover many of the things we learned in class."
- "We spent most of our time learning about one thing, and there was only one question on that topic."
- "I accidentally skipped over items I could answer because I didn't see them."
- "You don't give us enough space to write our answers."
- "The directions were confusing."
- "The questions are like you're trying to trick us."
- "It was hard to remember everything because we had two tests on the same day."
- "Sometimes I get so nervous and frustrated I give up."

Teachers were glad they'd asked. They carefully considered their students' feedback and used it to improve tests.

Students take many teacher-made tests throughout their school years. Educators use results from these tests to determine report card grades and honors, approve students for promotion and graduation, and monitor students' learning progress and the efficacy of instruction (Salend, 2009). However, as both research and experiences like those at Madison Middle School reveal, creating a good test is a challenge. Many students take poorly designed tests that negatively affect their performance and report card grades (Salend, 2011).

As a teacher and writer who focuses on assessment, I've worked with educators like the teaching team at Madison to help make tests more accurate and inclusive. I offer here

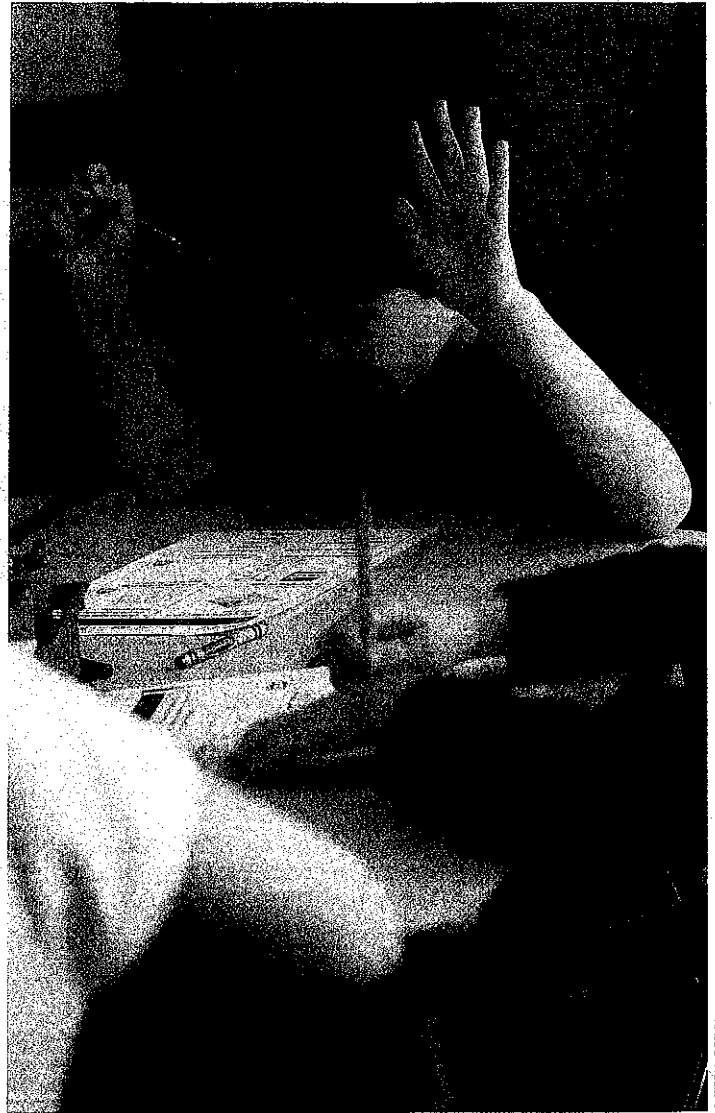


guidelines, strategies, and models for creating student-friendly tests—including before-and-after examples from Madison Middle School that show how teachers revised tests so that faulty test structures no longer hindered students from doing their best.

Fostering Validity

An essential aspect of creating student-friendly tests is ensuring *validity*—the extent to which the test measures what it

Friendly Tests



© STEFANIE FELIX

claims to measure. Invalid tests are unfair and of little value in helping teachers assess learning or determine fair grades.

Determine Scope and Weight of Test Items

Effective test creation begins with determining the scope of tests. A valid test should cover the main topics, concepts, and skills teachers taught during the time preceding the test. Valid tests also cover appropriate—rather than unrealistic—amounts of material.

“Tricky” questions that undermine a test’s validity may occur inadvertently if teachers use test questions that assess information in an entirely different way than they presented that information in class. In creating good test items, we should address not only what was taught but also *how* (Salend, 2009). Language used to present test directions and items must align with the terminology used during instructional activities. Essay questions are usually best for assessing content taught through role-plays, simulations, cooperative

learning, and problem solving; objective test items (such as multiple choice) tend to be more appropriate for assessing factual knowledge taught through teacher-directed activities (Salend, 2011).

In terms of weighting, the percentage, number, and point values teachers assign to test questions covering specific topics should directly relate to the difficulty of the content and the amount of class time devoted to teaching it (Salend, 2011). For instance, if 20 percent of instructional time was spent on teaching the events that led to World War II, then a corresponding percentage of test questions and point values should cover that topic.

Schedule with Sanity

As Madison students' comments indicate, the scheduling of tests can also affect students' performance. Frequent testing covering more specific content enables teachers to tell students what they should study, give students enough time to complete tests, and more accurately assess mastery. Good test scheduling means that teachers coordinate so students aren't overwhelmed with too many tests in one time period.

The Madison teachers began to administer regularly scheduled tests that assessed a reasonable amount of content rather than infrequent tests covering a great deal of content. They determined the content of their tests by identifying the most important topics and concepts they taught and the percentage of instructional time devoted to each. Teachers also collaborated to plan their testing schedules.

Fostering Accessibility

As the Madison team discovered, confusing tests hinder students' performance. These teachers drew on research and proven strategies to create tests

FIGURE 1 Sample Prompts for Tests

Prompting Students to Relax



Relax, take a deep breath



Smile, you are halfway through

Source: From *Creating Inclusive Classrooms: Effective and Reflective Practices* (7th ed., p. 472), by S. J. Salend, 2011, Upper Saddle River, NJ: Pearson. Copyright 2011 by Pearson. Reprinted with permission.

that were *accessible* by improving directions, format, readability, and legibility (Salend 2009, 2011).

Directions

Student-friendly tests have clear, complete directions that help students understand the context and conditions associated with items. Such directions say concisely what students are expected to do, note the precision students should provide in their answers (for example, angle measurements within a specific number of degrees), highlight point totals for items and sections, and present formulas and other

information needed to respond to questions (Salend, 2011).

Use language that students can understand and avoid vague terms (for example, *frequently*, *usually*) and irrelevant information that may confuse and frustrate test takers (Brookhart & Nitko, 2008; Elliott et al., 2010). Clear directions often feature the following:

- Numerals or number words to provide sequenced information in chronological order (Salend, 2009).
- Bullets to present crucial information that does not have a specific order (Rotter, 2006).
- Direction reminders (such as "Remember to write clearly and in complete sentences") throughout the test (Salend, 2009).
- Symbols that prompt students to pay attention to directions, such as color-coded arrows pointing to directions for specific item types (Elliott et al., 2010).

Format

It's important to set up test items in an organized way (Roach, Beddow, Kurz, Kettler, & Elliott, 2010). Presenting items in an intuitive, predictable, and numbered sequence helps students transition from one test question to the next and lessens the likelihood that they will skip items. Showing a reasonable number of items on each page, grouping similar question types together, and enclosing directions in text boxes all enhance student attention to test items.

Giving students an appropriate amount of space to write their answers helps them structure the length of their responses (Salend, 2009). It's best to have students record answers on the test rather than on a separate score sheet (Walker & Schmidt, 2004). Numbering test pages helps teachers give clearer directions and helps students locate—and ask questions about—specific items.

Readability

Examine and adjust the linguistic complexity of the text to make sure it's appropriate for students. To enhance readability,

- Eliminate unnecessary words.
- Reduce the length of sentences.
- Use language that's familiar to students and consistent with terms used in class.
- Employ a tense, sentence structure, and tone that students can understand. For example, "What is the perimeter of the figure?" is more readable than "What is the perimeter of the figure below, which comprises a square and an adjoining triangle?"
- Refer directly to important points, objects, or events rather than using pronouns.
- Avoid double negatives, ambiguous terms, abbreviations, contractions, acronyms, quotations, and parentheses (Salend, 2009, 2011).

There are several online resources that teachers can tap to assess a test's readability, such as the Test Accessibility and Modification Inventory (<http://peabody.vanderbilt.edu/Documents/pdf/LSI/TAMI.pdf>). In addition, most word-processing programs include readability formulas and strategies for enhancing a selection's readability.

Legibility

Good typographic and visual design choices increase a test's legibility and support students' understanding, clarity, and speed. I recommend the following choices:

Type size. Use 12- to 14-point type for most test takers and 18-point type (at least) for students with visual difficulties and beginning readers. Type that's too

small is difficult to read, and type that's too large causes the eye to make excessive movements.

Typefaces/fonts. Choose familiar typefaces or fonts (for example, Times New Roman) and avoid mixing fonts. Sans serif fonts (such as Arial) are preferable; they resemble hand lettering and so boost letter and word recognition. Avoid text in all capital letters.

Stylistic features for highlighting. Use stylistic features such as boldface and italics only to highlight brief parts of sentences (for example, key words) or to focus students' attention on specific sections. Italics and boldface are preferable to underlining, which can cause students to confuse letters (such as *y* and *v*). Draw attention to crucial aspects of tests, such as the directions,



by surrounding them with white space or placing them in boxes with thick, dark borders.

Line length. Because line lengths can affect reading fluency, present text in line lengths of approximately four inches. A four-inch line contains 7–12 words, assuming a 12-point font. When it's crucial to use more words to provide the context for understanding a question (for example, in sentence completion items), try to keep word clusters together on the same line.

Justification. Use left-justified or aligned text and staggered right margins. Avoid right-justified text, which causes uneven word and letter spacing and makes it harder to track the flow of text, and centered text, which slows reading (Salend, 2009, 2011).

Easing Anxiety and Fostering Engagement

Between 25 and 40 percent of students may experience high levels of anxiety that can interfere with their motivation, memory, attention, test-taking behaviors, and test performance (Cassady, 2010). Teachers can reduce this anxiety and help learners engage with—and possibly even enjoy—test-taking by giving clear directions, using prompts to support success, and providing students choices about test items.

Provide Prompts

Teachers can embed within tests phrases like "take a deep breath" and related images, such as a person in a yoga pose, to help test takers stay focused, calm, and motivated. Prompts like those shown in Figure 1 seeded throughout a test encourage students to pay attention, ask questions, maintain effort, and give themselves positive messages and

reinforcement as they proceed (Salend, 2011).

Provide Choices

Choice leads to more engaged test takers (Salend, 2011). If a test consists of 25 questions of varying types, for instance, an instructor might give students a choice to respond to any 20 items. When giving students options, it's important to identify the topics or items that students *avoided* and find alternative ways to assess mastery of this content.

Avoiding Trick Questions

As the Madison students' comments reveal, confusing test items can hinder student performance. Teachers can lessen the likelihood of this happening

by using well-written, grammatically correct, academically appropriate test items. Research has identified best practices for composing multiple-choice, true-false, sentence-completion, and essay items (Salend, 2009). The Madison teacher team tackled their problematic test items and revised those items to incorporate best practices.

Multiple Choice

Although multiple-choice items typically assess recall of important information, they can also assess students' *application* of content. In writing these types of items, the stem should provide the context for the answer and any relevant material and terminology, and it should contain only one major point. The item's answer alternatives should

all be viable choices that are shorter than the stem and that share common elements (such as the same grammatical structure and level of specificity). For example, if the correct answer is a poetic device, all choices should be poetic devices: Answer choices shouldn't include key words from the stem or categorical words like *always* that can tip off students to the correct answer.

In creating good test items, we should address not only what was taught but also how it was taught.

FIGURE 2. Original and Revised Multiple-Choice Item

Original Item:

Directions: Choose the best answer.

1. Which of the following choices presents the best definition of what the term *alliteration* means?
 - A) A poetic device where there is a comparison between two objects with the idea of providing a clearer meaning to one of them
 - B) A poetic device where the words used imitate sounds
 - C) A poetic device where inanimate objects are given human traits or abilities
 - D) A poetic device which involves the repetition of initial consonant sounds

Revised Item:

Directions: Circle the letter of the choice that best answers the question. Each multiple-choice item is worth 3 points.

1. A poet writes **Bertha blew big, blue bubbles**. What type of poetic device is the poet using?
 - a) Alliteration
 - b) Metaphor
 - c) Onomatopoeia
 - d) Personification

Source: From *Classroom Testing and Assessment for ALL Students*, by S. J. Salend, 2009, Thousand Oaks, CA: Corwin. Copyright 2009 by Corwin Press. Reprinted with permission.

To alleviate visual confusion, present answer choices vertically, ordered in a logical sequence. Highlight keywords in the stem, limit the number of choices to no more than four, and eliminate such choices as *all of the above* or *none of the above*. Figure 2 shows how the Madison teachers improved a multiple-choice question.

True-False

True-false items assess students' factual knowledge and understanding of specific concepts. However, many students have difficulty answering true-false items. These difficulties can be lessened if, for each item, teachers (1) present only one important point or relationship; (2) address material they have explicitly taught rather than information gained through intuition, common sense, or general knowledge; (3) use declarative statements that are clearly either true or false; (4) offer meaningful information and the context for responding to the question

(preferably one that interests students); and (5) highlight important parts of items. Write out the responses *true* and *false* so that students answer by circling one or the other, to avoid students confusing Ts and Fs under pressure.

True-false items should not contain double negatives. If items must be stated in the negative, highlight the negative words and phrases. Avoid vague terms that can mean different things to different students (like *usually*, *probably*, or *is useful for*); qualifying words that cue students that a statement is true (like *often*, *may*, or *usually*); and absolute words that hint that a statement may be false (like *all*, *entirely*, or *never*). Notice how the revised true-false item shown in Figure 3 is clearly worded, relates to an authentic situation, and is presented in an appealing context (being a fact checker for a website).

Sentence Completion

Sentence-completion items can be difficult because the information needed to complete the sentence often comes from print materials which, when taken out of context, may be vague. Make sure the statement provides a sufficient context for knowing what answer to provide. The question should address important information and have one clear answer, and the missing word or phrase must be meaningful. Keep word blanks to a minimum in each statement and locate the blank near the end of the statement. Researchers generally recommend that a one-word response, or a short phrase at most, should be enough to complete each sentence.

When creating these items, it's helpful to decide whether specific synonyms, abbreviations, misspellings, and other variations of the answer the teacher has in mind will be considered correct and to inform students of this in advance. Some teachers address the issue of multiple responses by offering

FIGURE 3 Original and Revised True-False Item

Original Item:

WRITE I OR E TO IDENTIFY WHETHER THE FOLLOWING STATEMENT IS TRUE OR FALSE.

1. It is not uncommon for the amount of solar radiation that reaches a location on the Earth's surface to vary based on the season.

Revised Item:

You have been hired by a website to tweet accurate facts about the sun and the moon. Identify whether the following statement should be tweeted or deleted by:

Circling **Tweet** if the statement is **true**.

Circling **Delete** if the statement is **false**.

The first one is done for you as an example.
Each true-false item is worth 1 point.

- Tweet** Delete 1. The amount of solar radiation that reaches a location on the Earth's surface varies based on the season.

a word bank of choices from which students can select to complete the statement. Make sure that the words provided share similar grammatical features (such as being similar parts of speech) and are presented in a logical order. Let students know if they can use words from the word bank more than once. (See www.ascd.org/ASCD/pdf/journals/ed_lead/el_201111_Salend_Examples.pdf for an example of a revised sentence-completion item.)

Essay Questions

Essay questions use either a restricted response format, which offers a structure to guide the content and format of responses ("How are stalactites and stalagmites both different and similar?") or an open-ended format, which allows greater flexibility in composing an answer ("Imagine you are a blogger

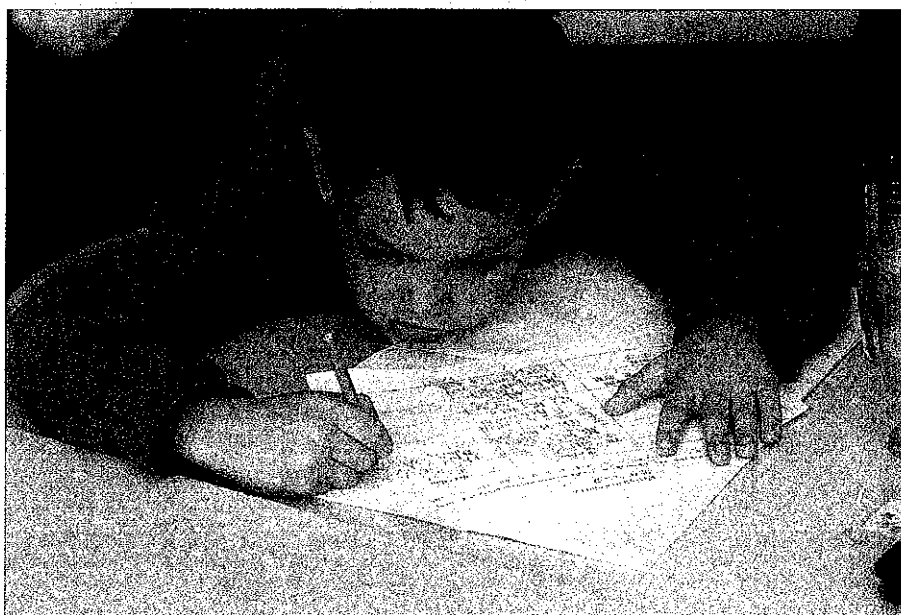
for your local newspaper. Write a blog entry about how electing the U.S. president by popular vote rather than by the electoral college would affect your community.").

Because of the numerous skills they demand, both types of essay questions present challenges for many students. Teachers can minimize challenges by making sure that their essay questions are focused and appropriate in terms of readability and level of difficulty. It helps to specify the essay's length and time limits, as well as what components should be included and what criteria the teacher will use to evaluate responses. Make sure to give students, especially those with writing difficulties, sufficient time to write their answers.

Teachers can help students interpret and answer essay questions correctly by

- Providing checklists of the

When we create vehicles that accurately assess student learning, we enhance the testing and grading experience.



components that should be included, to help students organize their responses.

- Dividing a larger, open-ended question into smaller, sequential subquestions.
- Displaying a list of important concepts that students should address in their essays. (See www.ascd.org/ASCD/pdf/journals/ed_lead/el_201111_Salend_Examples.pdf for an example of a revised essay item.)

Because essays measure students' skills in writing, higher-level thinking, creativity, and problem solving, rather than just factual recall, consider allowing students to use books and notes in writing their responses.

Ensuring Ongoing Improvement

After the Madison Middle School teachers revised their tests, they were pleased that their students' test performance improved significantly. They observed

that students were more motivated and less anxious when taking tests.

Like the Madison teachers, educators should continually evaluate their efforts to create student-friendly tests, primarily by examining whether students show improved test performance. They can check in with students about which questions they found difficult, easy, confusing, or frustrating—or simply why a student selected a particular response—and revise problematic items. Teachers should also ask students what surprised them about a test and what kinds of changes would improve that test or make students more comfortable while taking it.

I hope these suggestions and samples will help teachers create student-friendly tests that support better teaching and learning—and fairer grading. Teachers can also use these practices to improve premade tests they receive from

textbook publishers. When we create vehicles that accurately assess student learning, we enhance the testing and grading experience for everyone. ■

Author's note: This description of the effort to improve tests at Madison Middle School is a composite of assessment practices I've observed at various schools.

References

- Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Columbus, OH: Merrill/Pearson Education.
- Cassady, J. C. (2010). Test anxiety: Contemporary theories and implications for learning. In J. C. Cassady (Ed.), *Anxiety in schools: The causes, consequences, and solutions for academic anxieties* (pp. 7–26). New York: Peter Lang.
- Elliott, S. M., Kettler, R. J., Beddow, P. A., Kurz, A., Compton, E., McGrath, D., Bruen, C., Hinton, K., Palmer, P., Rodriguez, M. C., Bolt, D., & Roach, A. T. (2010). Effects of using modified items to students with persistent academic difficulties. *Exceptional Children, 76*, 475–495.
- Roach, A. T., Beddow, P. A., Kurz, A., Kettler, R. J., & Elliott, S. N. (2010). Incorporating student input in developing alternate assessments based on modified academic standards. *Exceptional Children, 77*, 61–80.
- Rotter, K. (2006). Creating instructional materials for all pupils. *Try COLA. Intervention in School and Clinic, 41*, 273–282.
- Salend, S. J. (2009). *Classroom testing and assessment for ALL students: Beyond standardization*. Thousand Oaks, CA: Corwin.
- Salend, S. J. (2011). *Creating inclusive classrooms: Effective and reflective practices* (7th ed.). Columbus, OH: Pearson Education.
- Walker, C., & Schmidt, E. (2004). *Smart tests: Teacher-made tests that help students learn*. Ontario, CN: Pembroke.

Spencer J. Salend is professor of educational studies at SUNY–New Paltz and author of *Creating Inclusive Classrooms: Effective and Reflective Practices* (Pearson, 2011) and *Classroom Testing and Assessment for ALL Students: Beyond Standardization* (Corwin, 2009); salends@newpaltz.edu.