

**Office of Assessment and Evaluation**

## **CCSD Assessment Development Guide**

## TABLE OF CONTENTS

PURPOSE.....	1
QUALITIES OF A GOOD ASSESSMENT.....	1
KINDS OF ASSESSMENTS.....	1
STEPS FOR ASSESSMENT DESIGN.....	3
Step 2: Consider the best format for measuring the identified standards/indicators.....	3
Step 3: Provide appropriate directions.....	6
Step 4: Ensure the readability level of the text is appropriate.....	6
STEPS FOR DESIGNING SPECIFIC KINDS OF ASSESSMENTS.....	7
MULTIPLE CHOICE ITEMS.....	7
MATCHING ITEMS.....	9
TRUE/FALSE ITEMS.....	10
FILL-IN-THE-BLANK/SHORT ANSWER.....	11
PERFORMANCE ASSESSMENTS.....	12
SOURCES.....	14

## PURPOSE

This purpose of this document is to provide a working definition of high quality assessments and to support the development of such assessments in Charleston County School District.

## QUALITIES OF A GOOD ASSESSMENT

**A good assessment is aligned with the standards** in terms of tested content and rigor. Thus, when developing an assessment, the first question you want to ask is “What is the best way to measure the desired content at the appropriate Depth of Knowledge (DOK) level?”

**A good assessment is as authentic as possible.** Authenticity refers to extent to which students demonstrate their learning in ways that would be applied in academic and professional settings.

**A good assessment is reliable.** Reliability concerns whether an assessment yields consistent results. In terms of a multiple-choice assessment, it is measured statistically by computing the extent to which each multiple-choice item gives you the same information about student learning. In terms of a performance assessment, reliability can be discussed in terms of the extent to which a teacher uses the rubric consistently or the extent to which different teachers interpret and apply the rubric in the same way

## KINDS OF ASSESSMENTS

Assessment falls into two categories: selected response and constructed response.

**Selected-Response Assessments** – require students to choose the best answer from a limited number of choices.

- Multiple-choice item - include a stem or question and four (4) response options, one of which is the correct answer.
- Matching questions - use two columns, one of which contains the stems, while the other contains the answers.
- True/false statements – ask students to identify if a statement is true or false.

**Constructed-Response Assessments** – require students to generate their own answers to a question or task.

- Fill-in-the-blank questions – provide an incomplete sentence, and a student is asked to generate the word or phrase that best completes it. A word bank is not provided. If a word bank is provided, fill-in-the-blank assessments items are considered selected-response assessment format.

- Short-answer questions - involve answering a question in a short phrase or two to three sentence response.
- Performance assessments – require students demonstrate their learning by producing a specific response or product. Examples include:
  - Oral performance, such as a debate, speaking in a foreign language, describing a picture, orally responding to an interview, reciting a poem, or giving an oral presentation.
  - Creation of a product, such as asking students to generate an essay, script for a play, lab report, painting, graph, journal, or poster.
  - Demonstrating a process, such as conducting a science experiment, playing a musical scale, finding a location on a map, tossing a basketball, or performing a dance.

## STEPS FOR ASSESSMENT DESIGN

### Step 1: Identify what you need to assess (standards and indicators) and develop a test blueprint.

A sample test blueprint is provided below. Note that standards/indicators are listed on the left, and the Depth of Knowledge (DOK) levels are listed across the top of the table.

The test writer should indicate within the cells of the table how many items will cover each standard/indicator at each DOK level.

#### Sample Test Blueprint

Standard/Indicator	DOK 1 Recall or Produce	DOK 2 Basic Application	DOK 3 Strategic Thinking	DOK 4 Extended Thinking	Total Number of Items/Weighting of Items
1					
2					
3					
4					
Total Number of Items					

To ensure the rigor of your assessment, review your blueprint to ensure that assessments are weighted to support learning in the course. For standards that contain a large amount of content, all content areas should be assessed.

### Step 2: Consider the best format for measuring the identified standards/indicators.

Some questions to help you determine the most appropriate format for measuring intended standards/indicators include:

- Which standards do you plan to assess, and at what DOK is each written?
- Are there constraints you should consider? For examples, how many students are you testing? How much time do you have for students to complete the assessment? What is your turnaround time for providing feedback?
- Are your students familiar with the format of the assessment you plan to administer? If not, you will need to familiarize them with the format prior to administering the assessment.

The table that follows identifies the advantages and limitations of various assessment formats.

### Advantages and Limitations of Various Assessment Formats

Assessment Format	Advantages	Limitations
Multiple-choice	<ul style="list-style-type: none"> <li>• Represents familiar format for students</li> <li>• Tests a large amount of material efficiently</li> <li>• Offers straightforward, more objective scoring</li> <li>• Can easily assess lower cognitive levels</li> </ul>	<ul style="list-style-type: none"> <li>• Test taker can interpret item differently; valid thinking may lead to incorrect answer</li> <li>• Students can guess and receive credit, even if they have no knowledge</li> <li>• Test-taking strategies may inflate scores of savvy test takers</li> <li>• If items only measure lower cognitive levels (DOKs 1 and 2), student memorization is encouraged</li> <li>• It takes substantial work (and may be impossible, depending on the standard/indicator) to create items that measure higher cognitive levels</li> <li>• Cannot be used to effectively assess students' ability to create an idea, product, etc.</li> </ul>
Matching	<ul style="list-style-type: none"> <li>• Useful for assessing content that includes many parallel concepts such as terms and definitions or causes and effects</li> <li>• Offers straightforward, more objective scoring</li> </ul>	<ul style="list-style-type: none"> <li>• Normally limited to lower cognitive levels, which may promote memorization</li> <li>• Frequently provide format clues to the answer; thus, test-taking strategies may inflate scores of savvy test takers</li> <li>• If students are required to select only one answer option for each stem/prompt, students may get one item wrong because they responded incorrectly to another item. Similarly, if they are able to correctly match one pair, it can help them correctly match another pair by eliminating an option, although they may not know the correct answer.</li> <li>• Students can guess and receive credit, even if they have no knowledge of the content</li> <li>• Cannot assess students' ability to create an idea, product, etc.</li> </ul>

Assessment Format	Advantages	Limitations
True/false	<ul style="list-style-type: none"> <li>• Can test a large amount of material efficiently; test-taker can respond quickly</li> <li>• Easy to generate true/false statements</li> <li>• Offers straightforward, more objective scoring</li> </ul>	<ul style="list-style-type: none"> <li>• Guessing is encouraged due to the likelihood of being correct without adequate knowledge</li> <li>• Test-taking strategies may inflate scores of savvy test takers</li> <li>• Normally limited to the knowledge level which may encourage students to believe oversimplified or ambiguous statements as truth</li> <li>• Cannot assess students' ability to create an idea, product, etc.</li> </ul> <p><b>NOTE:</b> <i>True/false items feel that the presentation of a false statement (negative suggestion) may have a detrimental impact on students' learning. Since the negative-suggestion effect would most likely apply to young children, true-false items should be used only with older students.</i></p>
Fill-in-the-blank & short answer	<ul style="list-style-type: none"> <li>• Minimizes guessing, requires demonstration of knowledge to receive credit</li> <li>• Promotes long-term retention better than selected-response items</li> <li>• Often more easy to construct than multiple-choice items when plausible distractors are difficult to generate</li> </ul>	<ul style="list-style-type: none"> <li>• Requires more time for the test-taker to answer than selected-response items</li> <li>• Scoring is often more subjective and time consuming</li> </ul>
Performance assessments	<ul style="list-style-type: none"> <li>• Can be a more authentic measure of student learning and hence may be more engaging way to assess students</li> <li>• Can assess highest cognitive levels as well as complex standards</li> <li>• Offers an opportunity for student learning</li> </ul>	<ul style="list-style-type: none"> <li>• Requires substantial time for completion</li> <li>• Grading is time consuming and more subjective</li> <li>• A more limited range of content/standards is addressed</li> </ul>

### **Step 3: Provide appropriate directions.**

All assessments should provide directions for students on how they should respond. For example, if students are asked to complete a matching exercise, it should be indicated if students may use the answers more than once.

All directions should be clear and concise. Directions should indicate all appropriate constraints. For example, are students allowed to use a calculator, collaborate with a peer, refer to notes, etc.? Time constraints and item weighting should also be noted.

### **Step 4: Ensure the readability level of the text is appropriate.**

Readability is a function of word difficulty, sentence complexity and the content of the material being assessed. You will want to ensure that the assessment is of an appropriate readability level for the grade and content for which it was designed.

Readability can be assessed using the following website: <https://readability-score.com/>. Readability level will be particularly important when you are measuring content outside of English Language Arts. Ideally, you want to reduce the readability level as much as possible while still ensuring that you are using appropriate vocabulary for the content area you are measuring.



# STEPS FOR DESIGNING SPECIFIC KINDS OF ASSESSMENTS

## MULTIPLE CHOICE ITEMS

Each multiple-choice item/question includes:

- an item stem – the question or statement that appears at the beginning of each item,
- four (4) response options –the options the student will choose between to select the correct answer, and
- three (3) distractors – the three (3) of the four (4) answer options that are NOT correct.

Some multiple-choice items may also include an interpretive exercise that could be a brief passage, table, figure, or map which the student must read or interpret in order to respond to the stem. Because interpretive exercises often take more time for students to process, ideally you should ask 3-5 questions per interpretive exercise.

## GUIDELINES FOR DESIGNING MULTIPLE CHOICE ITEMS

1. Provide directions for the student. This is especially important for interpretive exercises, which may be used to answer multiple items. In those cases, directions should indicate that the graph, table, etc. will be used to answer questions “X” through “Y.”
2. Have only one right answer.
3. Do not ask trick questions.
4. Try to make the item stem a complete sentence. If not, include the blank and ideally position it toward the end of the item stem.
5. Create plausible distractors. Words that students were exposed to in class but are not the correct answer as well as common errors that students make can be useful in developing plausible distractors.
6. Keep the length of the response options similar to avoid cueing the correct answer.
7. Avoid grammatical errors that may cue students to the correct answer. For example, the use of “an” before a blank indicates that the correct answer must begin with a vowel.
8. Ensure the question is not answered elsewhere on the test.
9. Arrange response options in a logical order (alphabetical, numerical/chronological) that does not cue the answer. Although some assessment development guides discuss ordering answer options by length, the correct answer is more likely to be the longest answer, resulting in the correct answer being “D” more often. Thus, ordering answer options alphabetically and also trying to make answer options similar in length should be considered.
10. Avoid using “none of the above” as a response option in conjunction with a negatively stated item stem—e.g., Which of the following is NOT an example of an animal?.
11. Avoid using “all,” “always,” or “never” in response options. These extreme words generally indicate that the answer is wrong.
12. Avoid using “usually,” “typically,” and “maybe” in response options. These tentative words generally indicate that an answer is correct.

13. Avoid using “all of the above.” If a student can eliminate one response option, then they know the answer also cannot be “all of the above,” so it artificially inflates their likelihood of getting the item right.
14. Avoid using “none of the above.” It can lead to students second guessing themselves unless the answer to the question is absolutely correct/without question, which is generally only the case in testing factual knowledge such as math questions which require computation or a history question that tests recall of a specific date.
15. Avoid lifting test items or phrases directly from students’ texts. When you use the same item or phrase to which students have already been exposed, you are more likely seeing if students can remember/recall rather than whether they have mastered content. In English Language Arts, this means using “cold read” texts and avoiding texts that are likely to have been read in class.
16. Make items as concise as possible and avoid repeated words in the response options. This is particularly important when you are measuring achievement OUTSIDE of English Language Arts. If items have a heavy reading load, students may not perform well if they have low achievement in ELA, even though they may possess the knowledge that you are trying to measure.
17. All items should have exactly 4 response options (1 of which is the correct answer).
18. Use numbers for the questions and letters for the response options.
19. Place each response option on its own line.
20. Punctuate correctly. When the item stem is an incomplete statement (e.g., includes a blank), options begin with a lower case (unless its proper nouns) and punctuation should be in the stem. When the question is a complete statement, options begin with a capital letter. If the options are complete sentences, then capitalize and punctuate as appropriate for each sentence.
21. Place easier items at the beginning of the test to lower student anxiety.
22. Review all test items and eliminate any sources of bias such as requiring students to have cultural knowledge to be able to respond to a mathematics item.
23. Ensure that you allow enough time for students to complete an assessment with multiple-choice items. Students can generally complete a multiple-choice item in about 1 -1.5 minutes.

## MATCHING ITEMS

Matching exercises use two columns: one column contains the stems or prompts, while the other column contains the answers (sometimes referred to as responses). The student is asked to match the stem with the correct. Matching items are most appropriate when you want to assess content that includes many parallel concepts such as the following:

- Terms and definitions
- Causes and effects
- Events and dates

## GUIDELINES FOR DESIGNING MATCHING ITEMS

1. Do not ask trick questions.
2. Every stem/prompt should have only one correct answer.
3. Use fairly homogeneous stems/prompts and answers (e.g., all names, all dates).
4. Provide directions to explain the format. This should include whether the answer options should only be used once or if they can be used more than once.
5. Have stems/prompts in one column with numbers and answers in another with letters.
6. Arrange items in a logical order, usually alphabetical or chronological.
7. All stems/prompts and answers should be visible on a single page.
8. Limit the number of stems/prompts and answers to no more than 10 per exercise so as not to overwhelm students.
9. Ensure your stems/prompts do not include grammar hints such as indicating the answer is plural, or vocabulary hints such as using the term itself in a definition.

## TRUE/FALSE ITEMS

True/false questions provide a statement which the students identify as true or false. When only two answers are plausible for an item, the true-false format is appropriate.

It is particularly challenging for teachers to write false items that are plausible enough to attract students who do not know the content. Another weakness of this type of item is that true/false questions are susceptible to error in measuring what students know and are able to do because of guessing. A student has a 50/50 chance of correctly guessing the answer to an item. Thus, these items should be used sparingly and ideally combined with other item types. If true/false items are included the following guidelines should be considered.

### GUIDELINES FOR DESIGNING TRUE/FALSE ITEMS

1. Statements should be absolutely true or false without qualifications. Do not mix partly true with partly false statements.
2. Have a balance between correct true and correct false responses. Since people who do not know an answer tend to use the true response, it can be beneficial to use slightly more false than true statements.
3. Have wording different from that of the text.
4. Avoid testing trivial information.
5. Avoid ambiguous language. Make the statements short, simple, and as clear as possible.
6. Avoid qualifiers as clues (e.g., all, none, only, could, might, may, sometimes, generally, some, few).
7. Avoid negatives, if possible, and always avoid double negatives. Some critics of true/false items feel that the presentation of a false statement (negative suggestion) may have a detrimental impact on students' learning. **NOTE:** *Since the negative-suggestion effect would most likely apply to young children, true/false items should be used with older students.*
8. Ensure that you allow enough time for students to complete an assessment with true/false statements. Students can typically respond to 1-2 true/false statements per minute.

## FILL-IN-THE-BLANK/SHORT ANSWER

Fill-in-the-blank items involve the completion of a statement, typically with a word or phrase. Fill-in-the-blank items are most appropriate when there are several correct responses to an item. If there is only one correct answer, a multiple-choice item is more appropriate. Short answer items are very similar but typically involve responding to a question in one to three sentences.

### **GUIDELINES FOR DESIGNING FILL-IN-THE-BLANK & SHORT ANSWER ITEMS**

1. Phrase the items so that the expectations for a response are clear and so that ambiguous answers cannot be used to cover for lack of knowledge. Stated differently, make it clear to students what information you are seeking. This may include formatting expectations. For example, for questions that require the student generate a number, identify the desired format of the number (e.g., should they provide units, should it be rounded to the tenth place). Structure the question to seek specific knowledge
2. Use fill-in-the-blank and short answer questions to test important content, not trivial details. For fill-in-the-blank, seek key words or phrases in the omitted part of the phrase.
3. If you are testing vocabulary knowledge, it is better practice (and assesses a higher cognitive level) to ask students to define the term rather than giving them the definition and asking them to provide the term.
4. For short answer items,
  - a. form a question that students respond to, rather than a statement, and
  - b. pose questions that can be briefly answered.
5. For fill-in-the-blank,
  - a. restrict the number of blanks to one or--at most—two, and
  - b. make the size of the blanks consistent with the desired response and ensure that the blank provides sufficient answer space.
6. Create a scoring key or checklist that provides all possible correct answers and identifies how many points to assign to partially correct answers. Align scoring key with the standard/indicator being assessed.
7. Score all student responses to each item at one time to ensure higher consistency in scoring.
8. Ensure that you allow enough time for students to complete an assessment with fill-in-the-blank items. Students can generally complete a fill-in-the blank item in about 1-1.5 minutes and a short answer item in 2 minutes.

## PERFORMANCE ASSESSMENTS

Performance assessments are often used to provide a more authentic assessment experience and because they can measure the highest cognitive levels. They require students to demonstrate their learning by producing specific responses or products. Performance assessments measure:

1. Oral performance – including debate, speaking in a foreign language, describing a picture, orally responding to an interview, reciting a poem, or giving an oral presentation
2. Creation of a product – including asking students to generate an essay, script for a play, lab report, painting, graph, journal, or poster
3. Demonstration of a process – including conducting a science experiment, playing a musical scale, finding a location on a map, tossing a basketball, or performing a dance.

## GUIDELINES FOR DESIGNING PERFORMANCE ASSESSMENTS

1. Determine the content or processes you want to measure based upon the content and process skills reflected in the standards.
2. Determine the context of the task. It may emerge from themes that were explored in the course, student backgrounds or interests.
3. Identify the audience for the performance (e.g. internal such as peers or other teachers vs. external such as parents or the public like a letter to the editor). Stated differently, in the context of a performance assessment in which students craft an essay, for whom should students write their essay?
4. Identify the task format and what students will be asked to do.
5. Based on completing steps 1-4, write the task directions. Develop the task directions in language that students can understand. Provide step-by-step directions. Within the directions, identify if the task will it be an in-class assessment, a take-home assessment, or some combination. If it is an in-class assessment, should it be completed in one class session or several sessions? Within the directions, explain whether students should work together, and if so, how (e.g., use peers as a “sounding board”/peer feedback vs. produce a group product).
6. Explain to students how they will be evaluated. This must include a rubric or proficiency scale that outlines the criteria used to evaluate the performance. Tools (rubrics or proficiency scales) that support performance assessments should be:
  - a. Aligned to standards and indicators. This means that if you are measuring science content knowledge, writing quality should not be heavily weighted in the rubric;
  - b. Consistent with the task directions;
  - c. Limited in scope so as not to overwhelm students;
  - d. Stated in terms of what the student does rather than what it does not do. Focus on the quality of student knowledge and skills when possible. Some verbs that can be used to focus on quality include accurate/inaccurate,

appropriate/irrelevant, clear/logical/confused, fluent/smooth/choppy,  
focused/broad, precise/vague, and thorough/limited/sufficient/superficial.

7. After developing a performance assessment, have a colleague and/or former student review the task and rubric or proficiency scale for clarity.
8. Revise the task based on student responses. You might also have a colleague score student work using the rubric or proficiency scale and identify whether it can be used consistently across raters or if it should be revised.

\*\*Note that for purposes of the discussion of performance assessments, it is understood that rubrics in lower grade levels may be an observation checklist.

## SOURCES

- Gredler, M., & Johnson, R. (2004). *Assessment in the literacy classroom*. Boston, MA: Allyn and Bacon.
- Gronlund, N. E. (2003). *Assessment of student achievement*. (7<sup>th</sup> Ed). Boston, MA: Allyn and Bacon.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. (3<sup>rd</sup> Ed.) Mahwah, NJ: Lawrence Erlbaum Associates.
- Haladyna, T.M., & Downing, S.M. (1989). A taxonomy of multiple-choice item writing rules. *Applied Measurement in Education*, 37-50.
- Hopkins, K.D., Stanley, J.C., and Hopkins, B.R. (1990). *Educational and psychological measurement and evaluation*. Englewood Cliffs, New Jersey: Prentice Hall.
- Kuhs, T., Johnson, R., Agruso, S., & Monrad, D. (2001). *Put to the test: Tools and techniques for classroom assessment*. Portsmouth, New Hampshire: Heinemann.
- Malamed, C. (2015). Tips for Writing Matching Format Test Items. Retrieved from [http://thelearningcoach.com/elearning\\_design/writing-matching-test-items/](http://thelearningcoach.com/elearning_design/writing-matching-test-items/)