

Assessment Methods—A Set of Four Options

Throughout our school careers, both as students and as teachers, we have encountered thousands of different assessments. Although the variations are endless, all of the assessments we have experienced and give today fall into one of four basic categories of methods:

1. Selected response
2. Written response
3. Performance assessment
4. Personal communication

All four methods are legitimate options, but only when their use is closely aligned with the kind of the learning target and the intended use of the information. (Portions of the discussion of methods are adapted from Stiggins, 2005.)

Selected Response

Selected response assessments are those in which students select the correct or best response from a list provided. Formats include the following:

- Multiple choice
- True/false
- Matching
- Fill-in-the-blank questions

Students' scores on selected response assessments are usually figured as the number or proportion of questions answered correctly.

Written Response

Written response assessments require students to construct an answer in response to a question or task rather than to select the answer from a list. They include *short-answer* items and *extended written response* items. Short-answer items call for a very brief response having one or a limited range of possible right answers. Extended written response items require a response that is at least several sentences in length. They generally have a greater number of possible correct answers.

Examples of short-answer items:

- Describe two differences between fruits and vegetables.
- List three causes of the Spanish-American War.
- What will happen if this compound is heated? Why will that happen?

Examples of extended written response items:

- Evaluate two solutions to an environmental problem. Which is better and why?
- What motivates (the lead character) in (a piece of literature)?
- Interpret polling data and defend your conclusions.
- Describe a scientific, mathematical, or economics process or principle.

We judge correctness or quality of written response items by applying one of two types of predetermined scoring criteria. One type gives points for specific pieces of information that are present. The other type takes the form of a rubric, which describes levels of quality for the intended answer.

Example of the “points” approach: When students in a biology class are asked to describe the Krebs cycle, points might be awarded for including the following information about the Krebs cycle:

- It describes the sequence of reactions by which cells generate energy
- It takes place in the mitochondria
- It consumes oxygen
- It produces carbon dioxide and water as waste products
- It converts ADP to energy-rich ATP

Example of the “rubric” approach: when students in an environmental science class are asked to evaluate two solutions to an environmental problem, their responses might be judged using these three dimensions: the criteria used for comparison, the accuracy of evidence brought to bear, and the strength of the argument for the supremacy of one over the other. (We cover rubric development in great detail in Chapter 7.)

Performance Assessment

Performance assessment is assessment based on observation and judgment. It has two parts: the task and the criteria for judging quality. Students complete a task—give a demonstration or create a product—and we evaluate it by judging the level of quality using a rubric.

Examples of demonstrations (skill targets) include the following:

- Playing a musical instrument
- Carrying out the steps in a scientific experiment
- Speaking a foreign language
- Reading aloud with fluency
- Repairing an engine
- Working productively in a group

Examples of products (product targets) include:

- Term paper
- Lab report
- Work of art

The rubric we use to judge the demonstration or product can award points for specific features that are present, or it can describe levels of quality. For example, to assess the ability to carry out a process such as threading a sewing machine, doing long division, or safely operating a band saw, points might be awarded for each step done correctly and in the correct order.

For more complex processes or products, you might have a rubric for judging quality that has several dimensions. In the case of evaluating an oral presentation, your rubric might cover content, organization, presentation, and use of language. Level of achievement will be reported by the number or percent of points earned, or in terms of a rubric score. (We address rubric development in depth in Chapter 7.)

Personal Communication

Gathering information about students through personal communication is just what it sounds like—we find out what students have learned through interacting with them.

Examples include the following:

- Asking questions during instruction
- Interviewing students in conferences
- Listening to students as they participate in class
- Giving examinations orally

Because these kinds of classroom assessments lead to immediate insights about student learning, they can reveal misunderstandings and trigger timely corrective action. This is why we usually think of them as formative, rather than summative assessments. As long as the learning target and criteria for judging response quality are clear, information gathered via personal communication can be used either way. Information gathered through personal communication can be used for instructional planning, as feedback to students to guide next steps, for student self-assessment and goal setting. But if the event is planned well and recorded systematically, the information can also be used as part of the final grade.

Student responses are evaluated in one of two ways. Sometimes the questions we ask require students to provide a simple, short answer, and all we're looking for is whether the answer is correct or incorrect. This is parallel to scoring for written selected response questions. Other times, our questions generate longer and more complex responses, parallel to extended written response questions. Just as with extended written response methodology, to evaluate the quality of oral responses we can use a rubric.

Matching Assessment Methods to Learning Targets

The accuracy of any classroom assessment turns on selecting the appropriate assessment method that matches the achievement target to be assessed. To begin thinking about the match between kind of learning target and assessment method, look at Figure 4.2

“Target–method Match.” The acceptable matches between methods and kind of learning target result in accurate information gathered as efficiently as possible. The mismatches occur when the assessment method is not capable of yielding accurate information about the learning target.

As you read through the Target-method Match chart, you will notice that the descriptions of the matches are described as *Strong*, *Good*, *Partial*, and *Poor*. Here is what each means.

Strong: The method works for **all** learning targets of this type.

Good: The method works for **many** of the learning targets of this type.

Partial: The method works **in some instances** for learning targets of this type.

Poor: The method **never** works for learning targets of this type

Assessing Knowledge Targets

Selected Response

This is a **good match** because selected response options do a good job of assessing mastery of discrete elements of knowledge, such as important history facts, spelling words, foreign language vocabulary, and parts of plants. These assessments are efficient in that we can administer large numbers of questions per unit of testing time and so can cover a lot of material relatively quickly. Thus, it is easy to obtain a sufficient sample of student knowledge from which to draw a confident conclusion about level of overall knowledge acquisition.

Written Response

Written response is a **strong match** for knowledge targets. It is especially useful for assessing blocks of knowledge and conceptual understanding, such as causes of environmental disasters, the carbon cycle in the atmosphere, how one mathematical formula can be derived from another, or the concept of checks and balances in government. Written response assessment is not as efficient as selected response in sampling broad domains of content because response time is longer. So, if time is fixed, the assessment will include few exercises. But the trade-off is the potential to get at deeper levels of knowledge and conceptual understanding.

Performance Assessment

Performance assessment is a **partial match** for assessing knowledge targets. First we'll consider when it can be a good match. Then we'll explore the potential problems.

If we pose a performance task that asks a student to rely on the knowledge and reasoning to display a skill or create a product that meets certain standards of quality and the student performs well, then we can draw the strong conclusion that the student was, in fact, a master of the prerequisite knowledge needed to be successful. In this case, the match between knowledge targets and performance assessment can be a strong one.

However, problems arise when the student performs poorly on the performance assessment. The key question is, why did the student not perform well? Was it due to the lack of prerequisite knowledge? Failure to reason well using that knowledge? If it was a demonstration-based performance assessment, was the problem inadequate skills? If it was a product-based performance assessment, was the poor performance due to a problem with creation of the product? We just don't know because all of that is confounded.

Here's an example. Let's say we assign a complex performance, such as writing and executing a computer program, and let's say our learning target is student mastery of prerequisite knowledge. When a student's program works well, we can conclude she possesses the prerequisite knowledge. The problem comes in when the program does not run successfully. Because of factors beyond the prerequisite knowledge that could have contributed to the failure, we can't know that lack of prerequisite knowledge is the reason for failure. We will have to do some follow-up probing to find out if the prerequisite

knowledge was there to start with. But, if our objective was to assess mastery of specific knowledge, why go through all this extra work? To save time and increase accuracy, just use selected response or written response assessments to assess mastery of the important knowledge.

Secondly, it is inefficient to rely on performance assessment to assess all content knowledge. A single performance task does require some subset of knowledge, and you can assess its presence with a particular performance task, but how many performance tasks would you have to create to cover all the knowledge you want students to acquire?

Thirdly, it isn't practical, or in some cases safe, to conduct certain performance assessments to assess knowledge. For example, if you want to assess students' ability to read bus schedules, although it would be most "authentic" to ask students to get around town on the bus, it would be highly inefficient and perhaps dangerous. Asking students to answer multiple-choice or short answer questions requiring understanding of a bus schedule would be a more efficient and safer way to get the information needed.

For these reasons we recommend as a general rule of thumb that you assess knowledge with a simpler method and reserve performance assessment for those learning targets that really require it.

Having said that, there are select situations in which performance assessment is a good choice for assessing knowledge. With primary students and with students who cannot read or write, we rely heavily on observation and judgment to determine mastery of knowledge targets. Selected response and written response are not good choices if students cannot yet read or write at a level that would allow them to show what they know.

Personal Communication

Personal communication is a **strong match** with knowledge targets for most students at all grade levels. While for summative uses it tends to be inefficient if a lot of knowledge is to be assessed, recorded, and reported for lots of students, it works well in formative applications, such as real-time sampling of student understanding during instruction. Additionally, for some students such as those with special needs, English language learners, or younger students, it may be the only way to gather accurate information.

Assessing Patterns of Reasoning

Selected Response

Selected response is a **good match** for reasoning targets. A common misunderstanding is that selected response questions can't assess reasoning proficiency. Although not a good choice for some patterns of reasoning, other patterns of reasoning can be assessed in selected response format. For example:

- Which of the following statements best describes how dogs in real life are different from the dog in the story? (Comparative reasoning)
- What generalization can you make from this selection about how these plants lure their prey? (Inference—generalizing)
- Which answer best explains the author's purpose in writing this story? (Inference—determining author's purpose)
- Choose the sentence that best tells what the story is about. (Inference—identifying main idea)

There are limits to this format when assessing reasoning. If you want to assess how well students can choose from their store of tactics to solve a problem requiring several steps, how well they can explain their choice or reasoning process, or how well they can defend an opinion, you must use another assessment method. For example, you might ask students to solve the following problem in mathematics: "Estimate the number of hours of TV advertising the typical U.S. fifth grader watches in a year. Describe the process you used to determine your answer." This is an extended written response question. If the learning target you want to assess falls into the category of student reasoning, a single number as the right answer is not the focus of the assessment—the reasoning process itself is.

Written Response

Written response represents a **strong match** for assessing reasoning targets. The trick here is to pose good questions, ones that require students to analyze, compare, contrast, synthesize, draw inferences, and to make an evaluative judgment. The criteria used to determine student scores must include the quality of each student's application of the pattern of reasoning in questions as well as the accuracy and appropriateness of the information or evidence brought to bear.

Also, remember from Chapter 3 that to assess a student's ability to reason productively, the question has to pose a novel problem to be solved at the time of the assessment. If students worked on the answer to the question during instruction, and that very question appears on a subsequent assessment, their answers are likely to represent a piece of remembered knowledge, which does not require reasoning.

Performance Assessment

This is a **partial match** for assessing reasoning targets, for the same reasons as with performance assessment and knowledge targets. We can observe students carrying out science laboratory procedures and draw strong conclusions about their reasoning based on our observations if they succeed at the performance assessment. However, if they

don't do well, it could be due to lack of prerequisite knowledge, lack of motivation, or to imprecise reasoning. Again, we confront the confounding of explanations. Without engaging in additional assessment, we remain unable to judge level of achievement on reasoning targets.

Personal Communication

For gathering accurate information, personal communication is a **strong match** to reasoning targets. Teachers can ask students questions to probe more deeply into a response. Or, students can demonstrate their solution to a problem, explaining their reasoning out loud as they go. The drawbacks with using personal communication to assess reasoning proficiency are the amount of time it takes and the record-keeping challenge it poses.

Assessing Skill Targets

Selected Response

Selected response is a **poor match** for skill targets. We can use it to determine if students possess the prerequisite knowledge required to perform skillfully, but it cannot be used to determine whether students can actually perform skillfully.

Written Response

Written response is also a **poor match** for skill targets, for the same reasons.

Performance Assessment

There is really only one assessment method that is a **strong match** for skill targets, and that is performance assessment. For example, we can determine whether they know how to conduct themselves during a job interview using another assessment method, but the only way to evaluate how well they can do it is to watch and listen to them during a simulated job interview and then judge their level of achievement.

Personal Communication

Personal communication is a **partial match** for assessing skill targets. It is a good choice when the skills in question fall into the category of oral proficiency, such as speaking a foreign language or giving an oral presentation. In these instances, personal communication is the focus of the performance assessment. When the skill target in question is not related to oral proficiency, such as “dribbles a basketball to keep it away from an opponent,” personal communication won’t do.

Assessing Proficiency in Creating Products

Selected Response

Selected response is a **poor match** for product targets. We can use it only to determine if students possess the prerequisite knowledge required to create the product, which is not the same as demonstrating the ability to create the product itself.

Written Response

Written response is a **poor match** for product targets. When the learning target specifies the creation of a written product, such as an essay or a research report, the appropriate assessment method is performance assessment. Remember, by definition, written response is a short or extended answer to a question or task, and by definition, we limit it to assessing knowledge and reasoning targets.

Performance Assessment

Performance assessment is a **strong match** for determining whether students can create a specified product: have them create the product and then judge its quality.

Personal Communication

Personal communication is a **poor match** for assessing product targets. We can use it only to determine if students possess the prerequisite knowledge required to create the product.