



Time: 18 minutes 19 seconds

Index

- Lesson Overview
- Video Outline
- Your Learning Objectives
- Your Stops
- Suggested Discussion Questions
- Recommended Activities
- Suggested Resources
- Script

Lesson Overview

In this lesson you'll be introduced to:

- what classroom summative assessment is;
- how classroom summative assessment compares to other types of assessment, including commercial and state assessments;
- characteristics of high-quality summative assessment, including reliability, validity, and use of various item types;
- suggestions for teachers and instructional leaders to improve classroom summative assessment practices; and,
- recommended activities and suggested resources (which are expanded on in this facilitator guide).

Video Outline

- Introduction (00:00-01:55)
- Classroom Summative Assessment Definition (01:56-03:36)
- High-Quality Classroom Summative Assessment (03:37-11:18)
- Classroom Summative Assessment vs Commercial and State Assessments (11:19-13:28)
- Improving Classroom Summative Assessment Practices (13:29-17:06)
 - Teachers
 - Instructional Leaders and Administrators
- Activities and Resources (17:07-18:19)
 - Recommended Activities
 - Suggested Resources

Your Learning Objectives

Record your objectives and points to focus on.

Your Stops

Make notes on stopping points and content discussion you would like the participants to take part in.

| Stopping Point | Content Discussion | Notes |
|----------------|--------------------|-------|
| | | |
| | | |
| | | |
| | | |
| | | |

Suggested Discussion Questions

- What are several ways in which classroom summative assessment differs from classroom formative assessment?
- What are the benefits of using “multiple measures” in the assessment of student learning?
- Explain how certain grading practices can affect students’ motivation to learn negatively.
- What is the relationship between test reliability and test validity?
- How are face validity and content validity alike and how are they different?
- Google “predictive validity.” Using a particular context or example, explain the relationship between predictive validity and content validity?
- How can advanced planning using a content-by-process matrix help assure the alignment of a comprehensive test to a set of content standards?
- What is meant by the test alignment category “range of knowledge”?
- In terms of domain coverage, why is a mix of item types important in the assessment of student learning?

Recommended Activities

- Find a test you or a colleague of yours administered to students after a significant unit of instruction or marking period. Categorize the items or test questions according to Webb's Depth of Knowledge levels. (Google "Depth of Knowledge" if you want more information on the levels.) What can you conclude about the test relative to DOK?
- Familiarize yourself with (or find information on) the domain of content and skills covered by the unit or marking period referenced above. Develop a content-by-process matrix for that domain; then assign the items in the corresponding test to their appropriate cells in the matrix. What can you conclude about the test based on your work above? Are there ways the test could have been improved? If so, how?
- Expand on the content within one of the content categories in your matrix by listing specific learning objectives or targets covered in the unit or marking period. (There should be a lot.) Then evaluate the range of knowledge in that content category covered by the test – i.e., how good a sampling of the objectives or learning targets do the test items in that category provide? How could that part of the test have been better?
- Read the article by William Schafer, paying particular attention to the section on grading practices. Determine whether any of your grading practices or your school's grading policies are ones that can inhibit learning. Talk to your supervisor or administrator about such practices.

Suggested Resources

Kahl, S. R. (2019). *Focusing More on Learning and Less on Test Scores*. Posted paper, Portsmouth, NH: RMC Research Corp. <https://www.assessmentworkshop.com/wp-content/uploads/2021/09/Learning-vs-Scores.pdf>

Schafer, W. D. (1993). *Assessment literacy for teachers. Theory into Practice*, 32(2), 118-126, Copyright 1993 College of Education, The Ohio State University. <https://doi.org/10.1080/00405849309543585>

Note: We recommend teachers acquire one of the two publications listed below, or a similar comprehensive resource for teachers, for use throughout their teaching careers.

Taylor, C. S. and Nolen, S. B. (2008). *Classroom Assessment: Supporting Teaching and Learning in Real Classrooms (2nd edition)*. Upper Saddle River, NJ: Pearson Education, Inc.

Nitko, A. J. and Brookhart, S. (2019). *Educational Assessment of Students (8th edition)*. New York, NY: Pearson Education, Inc.

ScriptSlide 1

- Welcome to Lesson 3 on classroom summative assessment.

Slide 2

- Much of the literature on balanced assessment systems discusses formative, interim, and summative assessment, with the latter two terms referring to *external* testing programs from test publishers or the state. The literature often ignores a very important category of assessment – classroom summative assessment.
- To students and their parents, results from classroom summative measures, and course grades based on them, are especially important. They are timely indicators of how students are doing in school and are used for decisions including the need for remediation or tutoring, advancement to the next grade in school, and graduation.
- Summative test results and course grades *matter*, and they should be fair and accurate.

Slide 3

- This lesson has four parts:
- One - What classroom summative assessment is and its role in a balanced assessment system
- Two - The attributes that assure high quality classroom summative assessment
- Three - Key elements of commercial and state assessments, and how they are similar to and different from those of classroom summative assessments
- And four - Actions you can take to improve classroom summative assessment practices and avoid some of the common shortcomings of many classroom summative measures.

Slide 4

- Why is this information important?
- It can help you draw more valid conclusions about student learning and make more appropriate decisions based on your classroom summative measures.
- It can help you assure fairness of classroom summative assessment and avoid unintentionally hindering student motivation to learn.

Slide 5

<blank>

Slide 6

- Recall from Lesson 1 that classroom summative assessment is what teachers use to assign or contribute to students' grades.
- In a balanced assessment system, classroom summative assessment is considered an internal assessment because, like formative assessment, it is mostly under the control of the classroom

teacher, who either creates or selects the tests or other measures he or she wants to use for grading purposes.

Slide 7

- You'll also recall from Lesson 1 that there is an important difference between classroom formative and summative assessment. That lesson discussed the importance of **not** counting toward grades student work used for formative assessment purposes.
- That practice is unfair and detrimental to student motivation to learn.
- Remember, summative assessment occurs **after the learning**.
- That does not imply that learning stops altogether, rather it is the recognition that there are reasons to make judgments about what or how well students have learned at the end of an instructional period – for example, at the end of a project, unit, or semester.

Slide 8

- Classroom summative assessment is distinguished from external interim or summative assessments which are typically mandated by the district or state.
- While these external assessments may sometimes contribute to decisions about students' needs for remediation or advancement, they are especially useful and well-suited for purposes of program evaluation.

Slide 9

<blank>

Slide 10

- High quality classroom assessments are reliable and valid.
- This means that they include enough items – for instance, a good sampling of the relevant content and skills -- and require students to apply both low-order and higher-order cognitive processes.
- In this part of the lesson, we'll explain how these attributes work together to create high quality classroom assessments.

Slide 11

- Look at these non-technical definitions of reliability and validity. *<pause>*
- They work pretty well for classroom summative tests for which the teacher is the ultimate judge of these test qualities.
- Notice “and measures it well” in the definition of validity. Those words refer to reliability. Thus, a test cannot be valid if it is not reliable. Reliability is a necessary but not sufficient requirement for validity.
- The additional requirement for validity is measuring the “right stuff.” Validity is closely tied to the intended use of the test results.

- Experts talk about many different kinds of test validity, but at their core is the common requirement that they measure the appropriate content and skills effectively.

Slide 12

- Suppose a new administrator wants to identify teachers with great classroom practices to serve as mentors for first-year teachers. This new administrator has no prior experience with the current teachers. She decides to visit each teachers' classroom for five minutes, and based on that, she picks the ones who should serve as mentors.
- Her "test" is an unreliable measure of the quality of classroom practices. If she had engaged in longer observations – such as observing teachers for several full class periods, she would have a much more reliable measure.
- However, even with the longer, more reliable measure, she may still have a validity problem. Perhaps she is observing the classrooms towards the end of the school year, when the teachers have virtually no management challenges, because they have already built good relationships with their students and established strong routines. To what extent has the administrator measured the "right stuff"?
- She might get a better idea of the quality of the teachers' classroom management skills by observing them early in the school year, when she could see how they developed those routines and relationships which resulted in classrooms with very few issues. Additionally, she would want to review teachers' lesson plans, tests, and other artifacts to assure their quality.

Slide 13

- Recall we said that adequacy of coverage of subject-matter is an important component of high-quality classroom summative assessments.
- Most teachers attend to the reliability and validity of their classroom summative assessments, even if they don't use those specific terms.
- A teacher wouldn't end a two-week instructional unit by giving a test that contains a single multiple-choice question. If a teacher wanted to measure students' proficiency with whole number computation, the teacher would be sure to include questions that involve many variations of computations pertaining to all four mathematical operations.
- In high quality classroom summative assessments, teachers would use *enough items* addressing a *good enough sampling* of the various learning targets or objectives covered in the relevant instructional units. They need to feel comfortable in the test results and any conclusions or decisions that he or she would base on them.

Slide 14

- A high-quality classroom summative assessment exhibits appropriate depth of knowledge by including a good mix of item types that tap into different levels of cognitive processing – in other words, low level skills as well as higher order thinking skills.
- This image illustrates the relationship between subject-matter domain coverage, and item types that tap into different levels of cognitive processing.
- In the top figure, we see that it's possible to achieve good coverage of subject matter if an assessment includes many multiple-choice items. However, multiple-choice questions mostly test isolated, low-level knowledge or skills.

- In the second figure, we see that we could achieve good coverage using fewer constructed-response items, which are the kinds of questions or tasks that require students to show their work or explain their thinking. Such items might be worth four or more points each. Answering them would involve multiple pieces of knowledge and skills, so fewer items would be needed to provide a decent sampling of the content domain.
- In the third figure, we consider extended performance tasks, which would include activities such as designing and conducting an experiment or completing a project and making a report or presentation. These are more time-consuming activities and provide evidence of deeper learning. However, the amount of time they require may mean too few can be used to provide for adequate domain coverage by themselves.
- Ideally, a combination of these three types of items or tasks would be desirable in high-quality classroom summative assessment, which could involve various measurement approaches within a single test or over time.
- Unlike state and commercial tests, which must be time and cost efficient and therefore tend to over rely on machine-scorable items, classroom assessments should not have those limitations.

Slide 15

- Again, high-quality classroom summative assessment would give appropriate attention to different levels of thinking or cognitive complexity.
- Webb's Depth of Knowledge levels provide a common framework for different kinds of cognitive processing.
- You may also be familiar with Bloom's Taxonomy, which is an older but similar framework.

Slide 16

- Take a look at the multiple-choice item and the constructed-response question in this slide. *<pause>*
- Notice the difference in the cognitive demand of the two items. *<pause>*

Slide 17

- Let's quickly recap what we've learned in Part 2.
- High-quality classroom summative assessments are reliable and valid. We can trust that they will consistently measure student knowledge, and that they are really measuring what we want the students to know. Reliable and valid assessments include enough items to accurately measure student knowledge, and the items cover a sufficiently broad sampling of the subject matter so that we know students have mastered the relevant objectives.
- Furthermore, they include multiple types of items (including constructed-response and performance tasks, which allow students to show higher levels of cognitive processing and deeper expertise in the content area).

Slide 18

<blank>

Slide 19

- In this section, we'll briefly touch on two ways in which commercial and state external assessments differ from classroom summative assessments.

Slide 20

- One of the problems with commercially available and state-mandated tests is that they are often dominated by multiple-choice questions, which can quickly and easily be scored by machines. These kinds of questions mostly tap into Webb's Level 1 and Level 2 thinking.
- Classroom summative assessment measures can be more flexible. Teachers can base their students' grades on a wider range of measures, to include not only different types of tests, but also project reports, oral presentations, and essays.

Slide 21

- As an educator, you may have heard about different types of validity in relation to state-mandated and commercially available tests, for which testing companies need to produce statistical evidence.
- While there are many types of validity, teachers most often rely on face validity: Does the test logically "appear" to cover the appropriate content and skills? Given that teachers are intimately familiar with the content of their instruction, face validity can be sufficient.
- Content validity is typically the result of a more rigorous design process that assures that the test covers the right content and skills.
- Testing companies developing state tests need to show that the tests meet rigorous specifications, which in essence, assure appropriate coverage of both content categories within a subject and cognitive process levels. This is actually assuring the alignment of tests to the state curriculum standards.
- To develop comprehensive teacher-made tests, such as semester or final exams, teachers should plan in advance to assure coverage of a good balance of content topics and a good balance of cognitive processes.

Slide 22

- Consider the following tips for teachers, coaches or administrators. You'll also find recommended activities and suggested resources in the Lesson 3 Supplement.

Slide 23

- Unfortunately, many classroom summative assessments emphasize only low-levels of student skills and knowledge. In some cases, this is partly due to state assessment systems which link teacher evaluation to student performance on state tests that measure lower-level skills.
- Additionally, many schools use online instructional tools that include tests that are scored by machines. Finally, teachers are very busy, and thus may have little time to develop and score more complex assessments of students work.

Slide 24

- However, there is no better evidence of student thinking than teachers' judgments of actual student work. Using good formative assessment practices, activating students as learning resources for each other, and using fewer summative tests can help ease some pressure on educators' time.

Slide 25

- In the lesson on formative assessment we explained the unfairness of counting toward grades work done for formative purposes. We also explained this practice's detrimental effect on student motivation to learn.
- There are many common grading practices that can inhibit learning.
- Generally, they send the message that the quality of work doesn't matter or that a bad test score can be offset by good behavior or completing assignments on time even if they are done poorly.

Slide 26

- Consider the following tips for teachers, coaches or administrators. You'll also find recommended activities and suggested resources in the Lesson 3 Supplement.

Slide 27

- For teachers . . . Learn to create and select high quality items and tasks. (Either of the two books listed in the Suggested Resources would be a valuable resource throughout your teaching career.)
- Every time you create a classroom summative measure, be mindful of subject matter coverage and depth of knowledge levels. For more substantial measures (such as marking period tests and semester exams), use a more systematic planning approach to assure balance in the coverage of both content topics and cognitive demand.
- Use items and tasks of different types in a test, and use different kinds of measures over time to tap both basic and higher-order cognitive skills.
- Avoid grading practices that inhibit student motivation to learn. The article by Schafer listed in the Suggested Resources is particularly relevant here.
- Classroom formative and summative assessment work together. Use good formative assessment practices and cut down on the amount of summative testing you do to produce course grades.

Slide 28

- For coaches and administrators . . . Periodically review your teachers' summative measures for quality, providing appropriate feedback for improvement.
- Be sure professional development experiences of your teachers are ongoing and collaborative, and include training in and monitoring of their classroom summative assessment practices.
- Adopt school-wide grading policies that lead to fair grading practices that do not inhibit student motivation to learn. (See the Schafer article listed in Suggested Resources.)

Slide 29

- These activities, in effect, ask you to evaluate aspects of content validity of a test of yours or a colleague.
- A third activity included in the Lesson 3 Supplement has you examining your and your school's grading practices, with a focus on practices that could possibly inhibit learning.

Slide 30

- The Kahl document deals with a common concern about an over-emphasis on test scores at the expense of learning. It also elaborates on the roles that various individuals can play in addressing this problem.
- The Schafer paper talks about standardized test score interpretation in reference to the kinds of scores that were reported three decades ago. However, the lessons are still relevant, and the section on grading practices that inhibit learning is priceless.
- Two books on student testing are cited also in the Lesson 3 Supplement. We recommend that teachers acquire one of them, or a similar comprehensive resource on testing, for use throughout their teaching careers.