

# 2

## Assessment Literacy 2.0

### In This Chapter You Will Learn:

- How and why assessment literacy can benefit all educators.
- The importance of formative assessment.
- Why a shared understanding of assessment-related terms is necessary.
- The differences between assessments *for* and *of* learning as applied to in-school CFAs, district benchmarks, and standardized achievement tests.

Every educator must understand the principles of sound assessment and must be able to apply those principles as a matter of routine in doing their work. Accurate assessment is not possible unless and until educators are given the opportunity to become assessment literate. [They] must understand student achievement expectations and how to transform those expectations into accurate assessment exercises and scoring procedures.

—National Education Association, 2003, p. 4

### THE NEED FOR ASSESSMENT LITERACY

Exactly what is **assessment literacy**? It is the ability to understand the different purposes and types of assessment in order to select the most appropriate type to meet a specific purpose. How can improved assessment literacy benefit educators and students?

When [teachers] know and understand principles of sound assessment, know how to translate those principles into sound assessments

and quality information about students, and because they involve students in the assessment process as part of their effective instruction, a range of benefits will accrue to all. (Stiggins, 1997, p. 7)

As educators develop and refine their own assessment literacy, they become more confident in their ability to make use of a greater variety of assessment tools in their assessment toolkit. As they learn the specific attributes of each type of assessment and gain experience creating and using each type, they can more effectively match the right tool to the right job. Learning how to design a variety of effective assessments, rather than over-relying on one particular type, educators become more adept at utilizing multiple measures to reveal student understanding.

Promoting a shared understanding of assessment literacy within a school faculty becomes especially important to grade- and course-level teachers when they design common formative assessments. Together the participating educators can deliberately select the particular type(s) of assessment that will best reveal their students' understanding of the unit learning intentions and student success criteria currently in focus. Should they later discover that the assessment type they first selected proves limited in providing the feedback they need on student progress, they will be better able to collaboratively decide upon a different type of assessment.

A fundamental principle of assessment literacy is rooted in the ability to answer the question, "Why assess?"

## THE RATIONALE FOR ASSESSMENT

Why do educators assess? There are a host of reasons. In general, educators want to know if, and to what degree, students are making progress toward mastery of particular concepts or skills in the standards. They use assessment results to determine levels of proficiency, to assign letter grades, and to communicate to parents where students are and where they need to improve.

For their more immediate purposes, educators assess students (1) to evaluate the effectiveness of instructional strategies and make adjustments as needed, (2) to give students feedback about what they currently know and can do, and (3) to show students how to use feedback to determine where they need to go next in their learning.

Educators assess student progress informally through ongoing observations, questioning, dialogue, and anecdotal note taking. When they need a more formal method, they select or design an appropriate assessment

matched to their intended purposes and then use the student results to answer their questions about student learning.

The bottom line as to why educators assess is to accurately determine their instructional impact on student learning. The feedback from students' assessment results provides the authentic evidence of the efficacy of educators' efforts.

W. James Popham (2003) asks educators to think deeply about the following four questions during the planning stages of instruction and assessment:

- What am I really trying to teach?
- What do my students need to know and be able to do?
- How can I translate the big curricular goals . . . into specific teachable components?
- What do my students already know about the topic I'm planning to teach? (p. 5)

Each of these questions is addressed within the ten CFA 2.0 design steps. Educators “unwrap” the unit Priority Standards to identify key concepts, skills, and corresponding levels of cognitive rigor in Step 2. This clarifies what they are “really trying to teach” and “what students need to know and be able to do.” By deciding the sequenced learning progressions in Step 9, they break down the “big curricular goals into specific teachable components.” The pre-CFA for the unit of study, created in Step 7, reveals students' prior knowledge, “what students already know about the topic.” From the resulting feedback, educators strive to correctly interpret student understanding in order to appropriately plan their instructional next steps.

No explanation about the purpose of assessments had a greater impact on my own understanding of assessment's essential function than these words of W. James Popham (2003):

Teachers use test [results] in order to make inferences about their students' cognitive status. Once those score-based inferences have been made, the teacher then reaches instructional decisions based (at least in part) on those inferences. *Educational assessment revolves around inference making.* (p. 60; italics added)

This passage again underscores the critical importance of *well-written* assessment questions. If the questions are flawed, the inferences educators derive from the students' feedback will be flawed, and the resulting instructional adjustments will likely not correct students' misunderstandings.

## FOUNDATIONS OF ASSESSMENT LITERACY

Assessment literacy is founded upon six sequential steps that inform instructional decision making, reprinted here from *Rigorous Curriculum Design* (Ainsworth, 2010):

1. **Know your purpose.** Determine exactly what it is you want to find out, what it is you want the assessment to do, and why you are administering the assessment in the first place.
2. **Determine the appropriate assessment that will accomplish your identified purpose.** In this context, “appropriate” means the specific type(s) or format(s) most likely to tell you what you want to know (selected response, constructed response, and performance based).
3. **Select or create a quality assessment.** Take great care in choosing questions from an external source and/or crafting the assessment questions yourselves. If a question is faulty in any way, and students answer it incorrectly, educators will later have to determine whether the question itself was the problem or whether students simply did not know the content upon which it was based. Decide whether the planned questions included in the assessment will enable you to make an *accurate* inference as to what students know and can do.
4. **Administer and score the assessment; analyze the assessment results.** Look for evidence of student learning, specific to your purpose, in the student responses. Conduct an item analysis, determining which questions individual students answered correctly and which ones they did not.
5. **Make an accurate inference.** This will be possible only if the assessment questions that you selected or created in Step 3 are of quality and provide valid and reliable data.
6. **Adjust instructional decisions in a timely manner.** Determine instructional next steps for students based on the inferences you have made. (pp. 137–138)

Because no single assessment can fulfill *all* the purposes of assessment or provide comprehensive evidence of student proficiency, Carol Ann Tomlinson (1995) states,

Fruitful assessment often poses the question, ‘What is an array of ways I can offer students to demonstrate their understanding and

skills?’ In this way, assessment becomes a part of teaching for success and a *way to extend rather than merely measure* learning. (italics added)

How then do these assessment *purposes* translate to definable assessment *practices*? Let’s start with a well-rounded understanding of what formative assessments are.

## FORMATIVE ASSESSMENT DEFINITIONS

There are many definitions and descriptions of **formative assessment**. Here is a sampling of definitions that collectively convey the essence of what formative assessment is and how it should be used:

- “Formative assessment is a planned process in which assessment-elicited evidence of students’ status is used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics” (Popham, 2008, p. 6).
- “Formative assessment is a loop: Students and teachers focus on a learning target, evaluate current student work against the target, act to move the work closer to the target, and repeat” (Brookhart & Nitko, 2007, p. 116).
- “The purpose of formative assessment is to provide feedback to teachers and students during the course of learning about the gap between students’ current and desired performance so that action can be taken to close the gap” (Heritage, 2008, p. 2).
- “Assessments *for* learning happen while learning is still underway. These are the assessments that we conduct throughout teaching and learning to diagnose student needs, plan for next steps in instruction, provide students with feedback they can use to improve the quality of their work, and help students see and feel in control of their journey to success” (Stiggins, Arter, Chappuis, & Chappuis, 2006, p. 31).
- “Assessment *for* learning is any practice which provides information to pupils about what to do to improve. Assessment *as* learning is any practice which takes the ‘what to improve’ into ‘how to improve’” (Clarke, 2008, p. 9).
- “An alternative to consider is ‘assessment as feedback.’ . . . As teachers derive feedback information from assessments that they set (for) their students, there can then be important adjustments to how they teach, how they consider what success looks like, how they recognize students’ strengths and gaps, and how they regard their own effects on students” (Hattie, 2012, pp. 125–126).

## DRAMATIC RESEARCH SUPPORT

Underscoring the practical reasons for formative assessment is an extensive body of supporting research. The influence of formative assessment associated with student learning gained widespread attention with the publication of the 1998 *Phi Delta Kappan* article by British researchers Paul Black and Dylan Wiliam (1998b) and their extensive review of the research on classroom assessment that same year. Their meta-analysis (1998a) concluded that student learning gains resulting from the use of formative assessment were among “the largest ever reported for educational interventions” (p. 61).

After summarizing a body of research on the impact of formative assessments distilled from four thousand studies spanning 40 years, Dylan Wiliam (2007–2008) concluded, “When well-implemented, formative assessments can effectively *double* the speed of student learning” (p. 36).

In *Rigorous Curriculum Design* (Ainsworth, 2010), I took the liberty of interpreting the opening words of Wiliam’s conclusion, *when well implemented*:

It is not enough just to *administer* formative assessments. To realize their full potential for improving student achievement, such assessments must be carefully constructed, student results must be thoughtfully analyzed, inferences must be accurately made, and subsequent instruction must be differentiated to meet student learning needs accordingly.” (p. 41)

Recently published educational research continues to underscore the effects that formative assessment, when effectively implemented, can have on raising student achievement levels. In *Visible Learning* (2009, pp. 2–3), world-renowned educational researcher John Hattie explains that any professional practice that can achieve a 0.40 effect size equates to *approximately one year of growth in student learning*. (See the glossary at the end of this book for a further description of **effect size**.)

Formative evaluation ranks fourth among all positive influences on student learning, producing an overall effect size of 0.90—equivalent to *more than two years* of student gains within a single academic school year. The effective use of feedback ranks tenth, with an effect size of 0.79 and a nearly similar result—*almost two years* of student growth.

Such impressive statistics provide a highly persuasive rationale in support of educators making formative assessment and feedback part of their bread-and-butter staples of educational best practice. Common formative assessments are a great way for educators and students to receive and utilize resulting feedback to correctly interpret student understanding and adjust instruction accordingly.

## KEY ASSESSMENT TERMINOLOGY

There are many terms associated with the word *assessment* and as many interpretations of what those terms mean. Often the terms become confusing for educators and result in misinterpretations of meaning that consume valuable collaboration time to clarify, time that could be more productively spent interpreting student learning needs and collaboratively deciding how best to instructionally meet them.

For example, educators often use interchangeably (and incorrectly) these assessment labels: formative, summative, interim, common, performance, diagnostic, progress monitoring, progress checks, and assessments *of, for, and as* learning, to name only a few.

Therefore, it is important to strive for a consensus of understanding about what frequently used terms mean, a key component of assessment literacy. It is important to create a printed lexicon or glossary of assessment terminology within a school and school system so that everyone can be on the same vocabulary page.

To help you begin this endeavor, specific terms that apply to the CFA 2.0 process are defined throughout the chapters and appear again in a comprehensive list at the end of the book. This glossary will help inform professional discussions and promote a school- and district-wide consensus of understanding. Let's clarify a few of those terms here.

## ASSESSMENTS *FOR* LEARNING: CLASSROOM AND COMMON

**Classroom formative assessments** include pretests or pre-assessments given to students before unit instruction occurs, informal checks to gauge student progress during instruction, and even a comprehensive assessment at the conclusion of the unit—if the results are used to inform instruction. Formative assessments are, by name and intention, formative. Thus, they are typically not used to assign grades. These assessments *for* learning yield diagnostic student feedback that educators use solely to inform and adjust instruction.

**Common formative assessments** are closely similar to classroom formative assessments with one exception: they are collaboratively designed by elementary grade-level and secondary course-level *teams* of educators who are all teaching the same unit of study to their students during the same timeframe. Common formative assessments include a pre-/post- design format, wherein students are given the same (or an alternate form of the same) assessment at the start of a unit and again at its conclusion. The

results of the pre-CFA help the team members determine students' prior knowledge and current understanding *before* instruction of a standards-based unit begins. The results of the post-CFA provide evidence of students' understanding *after* a unit of instruction has been concluded.

CFA teams collaboratively use assessment results to (1) accurately interpret student learning needs, (2) set individual classroom goals as well as grade- and course-level team goals for student improvement, (3) identify and share effective teaching strategies to accomplish these goals, (4) create appropriate lessons and activities for groups of learners or individual students, (5) plan ways to differentiate instruction and correct student misconceptions, and (6) inform students about their current progress so they can adjust their learning methods and strategies.

The National Education Association (NEA, 2003) explains why formative assessments *for* learning are so vital to students:

In the context of classroom assessment, however, one key purpose can be to use assessment results to *inform students about themselves*. That is, classroom assessments can inform students about the continuous improvements in their achievement and permit them to feel in control of that growth. Thus, classroom assessments become assessments *for* learning. Teachers involve their students in the classroom assessment process for the [express] purpose of increasing their achievement. (p. 6)

## ASSESSMENTS OF LEARNING: CLASSROOM AND COMMON

**Classroom summative assessments**, given by individual teachers, or **common summative assessments**, given by teacher teams, can occur at the end of a unit, quarter, trimester, semester, course, or an academic school year. Since these assessments take place after all instruction and student learning have ended, they are summative in both design and intent. They report the final results of student learning to the educators, to their students, to students' parents, and to administrators—typically to support the assignment of letter grades and/or levels of proficiency. Thus, they serve as assessments *of* learning.

When all instruction and related learning activities for particular standards have concluded, the results of summative assessments are not often used to improve student understanding for current students. Instead, teachers typically use these assessment results to judge the effectiveness of their teaching practices and to improve instruction of those standards for future students.

These dual purposes of assessment—formative and summative—are well expressed in the following statement from the NEA (2003): “Assessment must be seen as an *instructional tool* for use while learning is occurring and as an *accountability tool* to determine if learning has occurred” (p. 3; italics added).

## IS THIS ASSESSMENT FORMATIVE OR SUMMATIVE?

In *Visible Learning for Teachers*, John Hattie (2012) shares Bob Stake’s humorous maxim: “When the cook tastes the soup, it is formative; when the guests taste the soup, it is summative” (p. 144).

Confusion can exist in the minds of educators when attempting to classify an assessment given at the end of a unit as anything other than summative. One broad distinction is this: If the results from that assessment are not used to monitor and adjust instruction in order to improve students’ learning, the assessment is indeed summative. If those results are so used, the assessment can rightly be classified as formative.

Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (Black & Wiliam, 2009, p. 9)

Whether to regard an assessment as either formative or summative depends on the assessment’s purpose and how the assessment results are to be used. Here are three examples to illustrate:

1. If the assessment is simply a final measure of how students performed on multiple standards taught during the quarter, trimester, semester, or course of study, the assessment is obviously summative.
2. If an educator uses the results from an end-of-unit assessment in any way to inform instruction for the same students before or during the next unit of study, the results are being used formatively, even though the assessment itself is a summative measure used to determine student understanding of the unit learning intentions.
3. If an educator provides students with the opportunity to revise their work during the evaluation process and thus improve their performance on a particular assessment, the assessment should rightly be considered formative. After the students complete their revisions and the final evaluation is determined, the assessment can then be regarded as summative.

Whenever educators use the feedback results of any assessment in a diagnostic way—to correctly interpret student learning needs in order to instructionally meet those needs or to enable students to revise and improve their work—then that assessment is thought of as being formative. It is an *in-process* assessment.

CFAs, by name and purpose, are principally formative assessments, but they can serve a summative function after all of the formative uses of the assessment results are concluded at the end of a unit.

One important point of clarification: Even though educators (myself included) typically refer to an *assessment* as being either formative or summative, it is more accurate to say that educators are *using the results formatively* to adjust ongoing instruction or *using the results summatively* to measure the end of students' learning, rather than to assign those labels to the assessment itself.

Hattie (2012) emphasizes the distinction this way:

One major mistake is to consider that the notions of “formative” and “summative” have something to do with tests; in fact, there is no such thing called summative or formative tests. “Formative” and “summative” refer to *the time at which* a test is administered and, more importantly to the nature of the interpretations made from the tests. (p. 144)

## DISTRICT BENCHMARK ASSESSMENTS

It is a widespread practice for school systems to administer district benchmark assessments to elementary and middle school students at the beginning of the school year and at the end of nine-week marking periods or twelve-week trimesters throughout the rest of the year. The purpose of these assessments is to determine if students are on track for success on the annual large-scale standardized achievement tests. In some districts these assessments are designed to be formative, with results made available quickly so that educators can see how their students are doing and make instructional changes accordingly.

But more typically these assessments are summative, administered to survey student understanding of the grade- or course-level standards taught during an entire quarter or trimester. School and district administrators in particular use this data to see how students in each building and in all schools across the district are progressing prior to the annual standardized achievement tests. However, these assessments can come with a heavy accountability factor if and when the results are used to classify and rank educators, students, and entire schools as underperforming.

Classroom educators do not find benchmark assessments particularly useful unless they inform their current and ongoing instruction to help students improve. Robert Marzano (2010) states that benchmark assessments “frequently violate many of the basic assumptions underlying good formative assessment” (p. 9), and cites James McMillan in support of that assertion:

These tests, which are typically provided by the district or commercial test publishers, are administered on a regular basis to compare student achievement to “benchmarks” that indicate where student performance should be in relation to what is needed to do well on end-of-year high stakes tests. . . . Although the term *benchmark* is often used interchangeably with *formative* in the commercial testing market, there are important differences. Benchmark assessments are formal, structured tests that typically do not provide the level of detail needed for appropriate instructional correctives. (2007, pp. 2–3)

## POSITIVE WAYS TO USE BENCHMARK ASSESSMENTS

Even though periodic benchmark assessments place additional accountability demands on educators and increased testing demands on students, they can be beneficial *if* educators are able to use the results properly, that is, to improve instructional efficacy. A few of the positive ways educators can use benchmark assessments include

- *Intervening* appropriately for students who are far from proficiency well in advance of the large-scale summative assessments.
- *Accelerating* instruction effectively to help already-proficient students achieve advanced or exemplary levels of performance on the external assessments.
- *Modifying* existing assessments and creating alternative assessments similar to the format and rigor of large-scale assessments to assist English language learners and special needs students.

How closely these district benchmark assessments are aligned to state and provincial assessments in terms of assessment formats (selected response and constructed response), cognitive rigor, and vocabulary varies from one district to another. However, the more closely district assessments are aligned to large-scale assessments in terms of format and wording, the greater the likelihood that students will be familiar with the ways they are being expected to demonstrate their understanding on those external exams.

## LARGE-SCALE ASSESSMENTS OF LEARNING

The National Education Association refers to the annual assessments developed at the state level and then administered by local school districts as assessments *of learning*. In a report entitled *Balanced Assessment: The Key to Accountability and Improved Student Learning* (2003), the NEA pinpoints the essential purpose of such large-scale, external assessments:

When standardized tests are administered, they typically are intended to inform various policy-level and programmatic decision makers, as well as teachers, parents and the community, about student achievement. They are assessments *of learning*. Students are not the intended users. Rather, the tests inform others about students. (p. 6)

Large-scale assessments by themselves have minimal impact on an individual child's academic growth. The usual turnaround time it takes to receive results is a significant drawback that greatly limits the assessment's usefulness with regard to informing current instructional decision making. As a rule, in the United States it takes weeks and even months for schools to receive the student results of their annual state tests. By that time, students have moved on to the next grade, rendering useless that data to improve learning for those same students *unless* the next grade's teacher uses the data to identify the learning needs of those incoming students.

This is not to denigrate the administration of large-scale assessments, however. Analysis of large-scale assessment results can lead to broad changes in curriculum content, curricular sequencing, curriculum delivery, and enhancements of individual classroom test items (Sargent, 2004). Although all these changes can be very good, the actual data still will not give educators the specific and timely information they need to impact the learning of individual students they work with each day.

Ask educators, "What data about student achievement are most useful to you on a daily basis?" and their answer is almost always, "The data we collect from informal checks of student understanding—particularly when we use that feedback to determine where students currently are in relation to achieving the unit learning intentions."

Yet, paradoxically, the data that carry the greatest weight in terms of accountability are the data derived from students' performance on large-scale assessments. This is problematic because standardized achievement tests, even those that are aligned to content standards, are *instructionally*

*insensitive*. Here is how Popham (2013) defines that term and describes its relationship to external exams and teacher evaluations:

**Instructional sensitivity** is the degree to which students' performances on a test accurately reflects the quality of instruction specifically provided to promote students' mastery of what is being assessed. . . . If a standardized test is instructionally *insensitive*, it should have no role at all in evaluating the instructional ability of a teacher. (p. 63, bold and italics added)

## A BALANCED ASSESSMENT SYSTEM

Even though large-scale, standardized testing is likely here to stay, at best it can only provide snapshots of what a child knows and can demonstrate during the high-pressure weeks of spring testing each year. Looked at in isolation from other assessments, such on-demand snapshots are insufficient. To maximize their value, they should be presented as part of a “photo album” that shows evidence of student understanding acquired over time—an album that includes the results of formative and summative assessments along with student work products from authentic classroom performance tasks collected throughout a yearlong curriculum of multiple units of study. These “photos,” when viewed together as a whole, will show a complete picture of student growth in learning, even when the students have left for the next level of schooling and taken their photo album with them.

## DISCONTINUE MINIMAL-IMPACT ASSESSMENTS

As important as it is for educators to be on the same page of understanding regarding assessment terminology, it is even more important for them to take a hard look at the *usefulness* of each in-school assessment and to discontinue administering any assessments that are not yielding valuable information that can impact teaching and learning. This includes careful scrutiny of those assessments that are part of an adopted curricular program.

With so many different types of assessments educators are required to administer, they need to become very critical consumers in their selection of assessments. They can accomplish this by applying the first two steps of assessment literacy presented earlier in this chapter:

1. **Know your purpose.** Determine exactly what it is you want to find out, what it is you want the assessment to do, and why you are administering the assessment in the first place.

## 2. Determine the appropriate assessment that will accomplish your identified purpose.

Considering the amount of valuable time it takes educators to design, administer, score, and analyze assessment results each assessment needs to be worth that investment of time. If it is not contributing significantly to valid and reliable inferences about what students currently know and can do—and what they need next in their learning—you may want to give serious consideration to dropping it from the assessment roster.

Identify and focus on those assessments that truly have meaning and the potential for producing maximum impact on student learning. School leaders can greatly support classroom educators in this by encouraging them to rely upon their professional judgment and that of their colleagues to judiciously “weed the assessment garden.”

## CHAPTER SUCCESS CRITERIA

In the next chapter you will have a first look at the complete diagram of sequential steps for designing a quality CFA. As an ongoing reminder, this same diagram will reappear in each succeeding chapter with the particular step in focus highlighted. In this way you will be able to continually reorient yourself as to where you are in the entire CFA 2.0 process.

But first, please take a few moments and write your responses to the success criteria for this chapter, or simply evaluate your understanding of each statement on a scale of one (low) to five (high). If you give yourself a lower score, identify your learning gap and reread the related section(s) for clarification. As suggested at the conclusion of Chapter 1, if you are reading this book as part of a professional study group, share your thoughts and ideas with colleagues.

### Success Criteria:

- Explain how and why assessment literacy can benefit all educators.
- Summarize the importance of formative assessment.
- State why a shared understanding of key assessment-related terms is necessary.
- Describe the differences between assessments *for* and *of* learning as applied to in-school CFAs, district benchmarks, and standardized achievement tests.