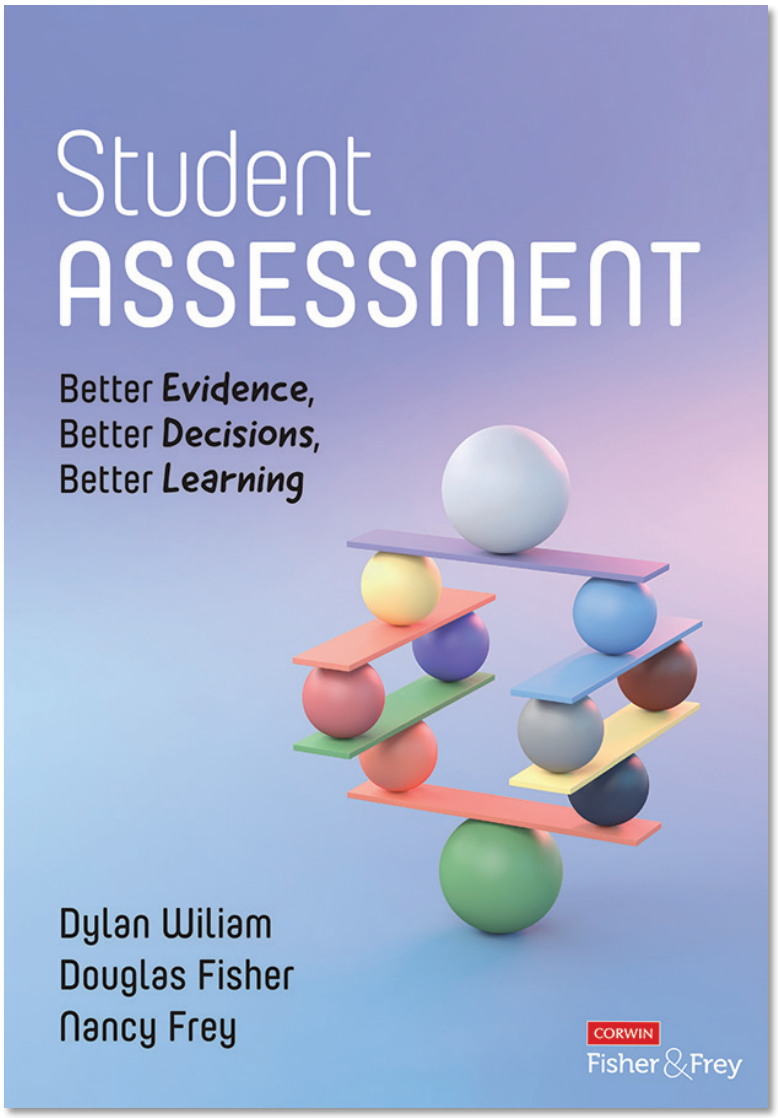

Thank you

FOR YOUR INTEREST IN CORWIN

Please enjoy this complimentary excerpt from *Student Assessment* by Dylan Wiliam, Douglas Fisher, and Nancy Frey.

LEARN MORE about this title!

CORWIN

Chapter

1

The Need to Assess

Assessment is an integral part of effective instruction due to both a principle about learning and an uncomfortable fact about the world. The principle about learning comes from David Ausubel (1968), who, over fifty years ago, in an introduction to a book on educational psychology, wrote this:

> If I had to reduce all of educational psychology to just one principle, I would say this: The most important single factor influencing learning is what the learner already knows. Ascertain this and teach [the student] accordingly. (p. vi)

The idea in this principle is simple: *Start from where our students are, rather than where we would like them to be.* Why is it so hard? Why do we need to assess? Because of the uncomfortable fact about the world: Our students often do not learn what we teach.

By this, we don't mean that our students *never* learn what we were teaching. Rather, our guilty secret as teachers is this: The sense that our students make of our instruction often bears little relationship to what we taught. That is why assessment is the bridge between teaching and learning. Only by assessing can we find out what sense our students made of our instruction.

Making assessment the bridge between teaching and learning involves two key shifts:

- A shift in how we think about aptitude
- A shift in how we think about teaching

Let's begin with a conversation about aptitude.

RETHINKING APTITUDE

For most of the last century, educators tended to define *aptitude* as the proportion of taught material that a student retained. Teachers taught a class, and while some students learned most or all of it, others retained much less. Assessments were generally conducted after many lessons or even weeks with little or no other evidence collected or used to guide learning. The idea that student achievement would approximate a normal distribution seemed natural and unproblematic.

However, in the late 1960s, this view was challenged by Benjamin Bloom. Drawing on the work of John Carroll, Bloom suggested a new way of thinking about aptitude: *the rate at which students learn with typical instruction.* While some students seem to learn quickly and others seem to learn more slowly, the rate at which students learn does not determine how much they *can* learn, provided they are given enough time and support. In fact, there is evidence that there are not "fast" or "slow" learners, but rather some students have had more experience than others and thus seem to learn faster (Koedinger et al., 2023). Returning to the 1960s, in Bloom's (1968) view, the fact that the results of students on a typical end-of-unit test resembled a "bell curve" was simply the result of ineffective teaching:

> In fact, we may even insist that our educational efforts have been unsuccessful to the extent to which our distribution of achievement approximates the normal distribution. (p. 3)

After all, if students differ in the rate at which they learn, then giving all students the same instructional experiences practically guarantees that we will get unequal outcomes. If, on the other hand, we think of aptitude as just the amount of time and support that students need to master material, then we can substantially reduce the range of achievement in a classroom by making the second shift in how we think about teaching: that it should be a *contingent*, rather than a *linear* process.

RETHINKING TEACHING

The idea that teaching should be a contingent—rather than a linear—process is hardly new. Over a hundred years ago, Frederic Burk (1913) criticized the idea of "lock-step schooling" and proposed, as an alternative, a system in which students in a class would progress at different rates according to their

capabilities. The "Individual System" (as it was known) led, in turn, to other individualized approaches to teaching, such as Helen Parkhurst's "Dalton Plan" (Parkhurst, 1922), the "Winnetka Plan" (Washburne, 1941), and the "Kent Mathematics Project" (Banks, 1975).

The distinctive feature of each of these approaches to instruction was this: What the student would do at the conclusion of an instructional activity would be determined only after the impact of those activities had been established. Instead of thinking of teaching as a linear process, where the next steps were determined solely by what had already been covered, innovative educators began thinking about teaching as a *contingent* process, using evidence about the effects of previous instruction to determine what should happen next. In other words, rather than simply using assessments to determine the effects of instruction once the instruction had been completed, teachers also began using assessment to *improve* instruction.

USING ASSESSMENT TO IMPROVE LEARNING

The first use of the phrase "assessment for learning" appears to be in a book published in 1973 titled *Assessment for Learning in the Mentally Handicapped* (Mittler, 1973). In retrospect, this is hardly surprising. Special education has always taken the approach of identifying individual learning needs through assessment, and it is worth noting that the first systematic review of studies on formative assessment was also conducted in special education (Fuchs & Fuchs, 1986). Since then, the phrase "assessment for learning" has been popularized through the work of Richard Stiggins and the Assessment Training Institute in North America (see, for example, Stiggins et al., 2004) and the Assessment Reform Group in the United Kingdom (Assessment Reform Group, 2002), as well as a number of other authors.

While Stiggins (2005) himself suggested that assessment for learning represented a particular approach to formative assessment, many other authors have used the terms "assessment for learning" and "formative assessment" interchangeably. This has caused some confusion, because there are many ways that assessment can improve learning, and grouping them all together results in a lack of focus that can hinder effective implementation. Let's look at some of the distinctions between them.

FROM ASSESSMENT OF LEARNING TO FORMATIVE ASSESSMENT

In teasing out the differences between assessment for learning and formative assessment, the first thing to note is that the phrase "assessment for learning" is a statement about the *purpose* of assessment, rather than the role it actually serves, as the following definition makes clear.

> The phrase "assessment for learning" is a statement about the *purpose* of assessment, rather than the role it actually serves.

> Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence. An assessment activity can help learning if it provides information that teachers, and their students, can use as feedback in assessing themselves and one another, and in modifying the teaching and learning activities in which they are engaged. Such assessment becomes "formative assessment" when the evidence is actually used to adapt the teaching work to meet learning needs. (Black et al., 2004, p. 10)

As Benjamin Bloom noted many years ago, one way that assessment can promote learning is by motivating students to study when they otherwise might not have done so. While many people have criticized the use of tests for motivation (see, for example, Kohn, 1999)—the fact remains that such things as quizzes and tests can get students to study more than they would do otherwise. Whether that study is productive or not is a different issue, of course, but it is fairly clear that the presence of a formal assessment of some kind does increase student achievement (Crooks, 1988; Natriello, 1987; Wiliam, 2010).

When educators put assessments in place to motivate students to study, it would be fair, then, to describe the use of such assessments as assessment for learning. Indeed, the first widespread use of the phrase "assessment for learning" was proposed by the Assessment Reform Group in the United Kingdom, who saw the use of portfolios and other forms of authentic assessment as a way of making secondary schooling more engaging to students (Broadfoot et al., 1999). The use of assessment processes to motivate students *would* therefore count as assessment for learning but would not necessarily be formative assessment.

A second way in which assessment can improve learning is by giving students the opportunity for what is commonly called "retrieval practice." It is common, when discussing educational assessment, to remark that "weighing the pig doesn't fatten it," but while this might be true in agriculture, it is wide of the mark in psychology.

In a review of different strategies that students might use to improve their learning, John Dunlosky and colleagues (Dunlosky et al., 2013) found that practice testing (self-testing or taking practice tests over to-be-learned material) was more effective than the strategies that students typically used (rereading, writing summaries, highlighting, etc.). Indeed, the effectiveness of practice testing is one of the most solidly grounded findings in all of cognitive psychology (Adesope et al., 2017; Carpenter et al., 2022).

The effectiveness of practice testing is one of the most solidly grounded findings in all of cognitive psychology.

Many educators and learners believe it's counterintuitive or even implausible that practice testing of the material that students need to learn can be more effective than rereading it. However, the finding is less surprising in the light of recent research on how human memory works, and in particular the work of Elizabeth and Robert Bjork, which we'll discuss shortly.

General understandings of how memory works seems to be similar to that proposed by Edward Thorndike (1913) over a hundred years ago: If things in memory are routinely used, then they are easy to recall (what Thorndike calls the "law of use"), and if things are not routinely used, then they become harder to recall (the "law of disuse"). However, in the "new theory of disuse" the Bjorks (1992) suggest that any item in memory has two characteristics:

- **Storage strength**: How well an item has been learned at any point in the past
- **Retrieval strength**: How easy an item is to recall right now

Retrieval strength goes up and down—things that used to be easy to recall can become harder to recall—but storage strength, being a measure of how strongly connected something is to other items in memory, can only increase (unless there is brain damage).

According to this theory, rereading a passage increases both storage strength and retrieval strength, but retrieving something from memory increases storage strength and retrieval strength even more. Further, the harder it is to retrieve things from memory, the greater the impact its successful retrieval has on long-term memory (Bjork & Bjork, 1992).

In this way, the "new theory of disuse" explains why practice testing is so effective. Rereading things improves retrieval strength, so if learners are tested on what they have read immediately, then they are likely to do well. But if the goal is to remember things for the longer term, then retrieving things from memory is better.

The conclusion here is that taking practice tests on the things students want to learn—anything from flashcards to more formal assessments—improves learning, *even if the tests are not scored*. The main value of practice testing is in providing practice in retrieving things from memory. For this reason, it would be appropriate to call practice testing a kind of assessment for learning—after all, it is an assessment administered for the sole purpose of improving learning—but it is not necessarily formative assessment, in that the assessment does not really *form* the direction of future learning.

> The main value of practice testing is in providing practice in retrieving things from memory.

If the tests are, in fact, scored, then there is another potential benefit, which is the result of a recently discovered psychological phenomenon known as the *hypercorrection effect*. In one experiment, Brady Butterfield and Janet Metcalfe (2001) asked undergraduate students a series of general knowledge questions. After each question, the students rated how confident they were that their answer was correct on a seven-point scale (–3 to +3). After each question, the students were told whether their answer was correct or not. If the answer was incorrect, the students were shown the correct answer for two seconds. The questioning continued until each student had answered fifteen questions correctly and fifteen questions incorrectly. Then, after a period in which the students did an unrelated task, they were retested on thirty questions: fifteen they had answered correctly and fifteen they had answered incorrectly.

As might be expected, there was a positive correlation between confidence and success—higher confidence was associated with higher accuracy. The researchers also found that there was a tendency for students to repeat

errors made with high confidence (again, perhaps not surprising). What was less obvious was that the researchers found that students were more likely to correct errors made with high confidence than those made with low confidence (Butterfield & Metcalfe, 2001).

A more recent analysis looked at the performance of seven thousand middle-school students who answered a question on a math test, and provided a rating of how sure they were that the answer was correct. A few weeks later, they completed a parallel test question. When the students answered incorrectly and indicated little confidence in their answer (1, 2, or 3 on a five-point scale), they answered the parallel question correctly 40 percent of the time a few weeks later. When they indicated high confidence in their incorrect answer (5 on a five-point scale), they answered the parallel question on the second test correctly 50 percent of the time (Foster et al., 2022). It is also worth noting that while some confident students go back to incorrect answers in later tests, most do not, and the benefits of the hypercorrection effect also extend beyond rote memorization (Corral & Carpenter, 2022).

The hypercorrection effect therefore provides a way in which assessment can improve learning. When students find out that answers they were confident are correct in fact are incorrect, there can be a substantial increase in their learning. Making use of the hypercorrection effect in instruction would then be an example of assessment for learning, because the assessment is being used to improve learning. However, whether it is an example of formative assessment is less clear. The learner is being told the answer is incorrect, and that might make the student ready to learn the correct answer, but the assessment itself is not forming the direction of future learning. As the earlier quote from Black et al. (2004) indicated, assessment for learning becomes formative assessment *when the information generated by the assessment is used to adapt instruction to better meet student needs.*

These ideas are summarized in Figure 1.1. Announcing that there will be an assessment of some kind is likely to motivate at least some students to prepare for the test, so we have an example of assessment for motivation. If the assessment is actually administered, then the students taking the assessment get the benefit of retrieving things from memory, thus making the memory stronger. If the students are told which of their answers are correct or not, then this may result in enhanced learning via the hypercorrection effect. However, to be formative—for the assessment to form the direction of future learning—the information from the assessment has to be used.

Figure 1.1 • Transforming Assessment for Learning to Formative Assessment

Announced?	Given?	Scored?	Used?	
✓				Assessment for motivation
	✓			Retrieval practice
		✓		Instructional correctives
			✓	Formative assessment

One objection that is sometimes raised at this point is that distinguishing between assessment for learning and formative assessment is an academic exercise that is of more interest to researchers than to teachers. However, we think that drawing out the different ways that assessment can improve learning has two significant benefits:

1. It allows us to ensure that people are not talking at cross purposes. If some educators use a term like *assessment* for learning to describe the use of tests to motivate students to study, while others use the term to describe how teachers and students can fine-tune their next instructional steps to maximize learning, then they are unlikely to have productive discussions.
2. The different mechanisms by which assessment can improve learning operate in different ways, drawing on different research bases. Being clear what precisely is being discussed is essential if we want to move on from *What works?* to *How much will it improve learning, and under what circumstances?*

Now that we have clarified the relationship between assessment for learning and formative assessment, we think it will be helpful to identify exactly what we mean by formative assessment.

FORMATIVE ASSESSMENT DEFINED

Over the last fifty years, many definitions of *formative assessment* have been proposed. Some researchers, such as Benjamin Bloom, suggested that formative evaluation (as he called it) involved the use of short tests during periods of instruction:

> Quite in contrast [to summative evaluation] is the use of "formative evaluation" to provide feedback and correctives at each stage in

> the teaching-learning process. By formative evaluation we mean evaluation by brief tests used by teachers and students as aids in the learning process. While such tests may be graded and used as part of the judging and classificatory function of evaluation, we see much more effective use of formative evaluation if it is separated from the grading process and used primarily as an aid to teaching. (Bloom, 1969, pp. 47–48)

Others, such as Richard DuFour (2007), envisaged formative assessment as a more formal process, with "common formative assessments" being administered to all the students in a particular grade at intervals of six to ten weeks, providing evidence about the students' progress toward mastery of the relevant standards. Dylan, in collaboration with Paul Black, Christine Harrison, Clare Lee, and Bethan Marshall, also saw the regular "checks for understanding" that teachers undertook in their teaching activities as a process of formative assessment—not least because thinking of checking for understanding as an assessment process draws attention to the quality of the evidence that teachers have on hand for the instructional decisions they need to take (Black et al., 2003). In response to the range of definitions proposed for formative assessment, some authors, such as Lorrie Shepard (2008) and W. James Popham (2006), suggested that the kinds of formative assessment processes proposed by Richard DuFour and others should not be called formative assessment since they were so different from the approaches from which the evidence of effectiveness was derived. To their thinking, the term should be reserved for shorter time cycles, rather than those administered at the quarter level. Overall, the conversation in the field grew from definitions of what assessments look like to considerations about when they are administered.

However reasonable such arguments might be, this has not stopped people from claiming a wide range of practices—from frequent checks for understanding, all the way to interim and benchmark tests—as being formative assessment. Rather than getting into these "turf wars," therefore, we think it makes sense to adopt an inclusive (and perhaps literal) definition of *formative assessment*, based on the extent to which evidence from the assessment *forms* the direction of learning. Drawing on the work of Black and Wiliam (2009), we suggest the following definition of formative assessment:

> An assessment functions formatively to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, students, 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 that elicited evidence.

Several features of this definition are worth drawing out in detail.

1. *The definition focuses on the function that the evidence from the assessment serves rather than the assessment itself.* The reason for this is that any assessment can be used formatively or summatively. For instance, if we give a third-grade student a test of twenty multiplication facts chosen at random from 1 x 1 up to 10 x 10, and the student gets 10 of them correct, then because we have chosen them at random, we can conclude that the learner knows approximately 50 percent of the number facts. This is *a summative* conclusion. If, on the other hand, we notice that the student appears to be having difficulty with the "seven times" table, then this gives us something to work with. This is *a formative* conclusion. Note here that the same assessment, and even the same assessment evidence, can be used summatively or formatively, so it makes little sense to talk about "a formative assessment" or "a summative assessment." It is the conclusions that we draw, rather than the assessments themselves, that are formative or summative.

2. *The information elicited by the assessment can be used by teachers, students, or their peers.* This is important because early definitions of "assessment for learning" focused on the role of the teacher, necessitating the invention of another term, *assessment as learning*, to describe the role of the learner (Earl, 2003). While the idea that students should be learning something while being assessed is attractive, equating assessment with learning is potentially unhelpful, since the term *learning,* at least in psychology, is used to describe a relatively long-term change in what students know, understand, or can do, while an assessment is basically a procedure for drawing conclusions (Cronbach, 1971).

3. *The definition focuses on decisions, rather than intentions or actions.* If an assessment was intended to elicit evidence to improve instruction but the evidence was not, for some reason, actually used, then the assessment would not be functioning formatively. This of course is similar to the issue with the term *assessment for learning*—a statement of intent rather than function—discussed earlier. Some definitions of formative assessment have focused on the effect of the assessment, but given the complexity of human learning, it seems likely that even the best formative assessment processes may occasionally be ineffective. Focusing the definition on what is *likely* rather than on what is certain makes for a definition that is actually useful.

4. *The definition allows for the situation in which the evidence does not change the decisions but merely confirms that the intended action is the best one.* For example, this occurs when a teacher checks on a class's understanding by asking all students to write answers on personal whiteboards and then, seeing that all the students have answered correctly, decides to move on.

Given the inclusive nature of this definition, we can then classify different approaches to formative assessment in terms of the length of the cycle and what is formed by the formative assessment. We will provide some brief explanations here and then discuss these assessments in more detail in the following chapter.

Long-cycle formative assessment *involves cycle lengths of four weeks or more—typically six to ten weeks—and can improve student achievement by monitoring.* This approach ensures that any students who are not making the progress needed to reach mastery of the applicable standards by the end of the year are identified, and appropriate action is taken (see e.g., Saunders et al., 2009). Long-cycle formative assessment can also help teachers ensure that the curriculum is aligned to the standards in place (Goe & Bridgeman, 2006) by relating student achievement to the curriculum.

Medium-cycle formative assessment *typically occurs within an instructional unit.* It can take the form of brief tests, as envisaged by Benjamin Bloom in the earlier quote, but it can also involve making students more active participants in the assessment process, so that assessment becomes something that is done *with* students rather than *to* them (Stiggins, 2001). An example would be making sure that students understand the criteria against which their work will be assessed.

Short-cycle **formative assessment** *occurs within and between lessons, day-to-day and even minute-to-minute.* This occurs not so much every six to ten weeks, but rather every six to ten minutes!

Conclusion

The answer to the question posed near the beginning of this chapter—Why assess?—should now be clear. We assess to make our teaching more responsive to our students' learning needs, to increase student engagement, and to strengthen our students' memories.

(Continued)

(Continued)

If we reframe aptitude not as "amount of material retained" but as "time needed to learn," then it becomes clear that if we make teaching a contingent process rather than a linear process, then many (and perhaps even most) of our students can reach not just proficiency but also advanced levels of achievement. A recent study, looking at the learning trajectories of almost seven thousand students from elementary grades to college in math, science, and language, found that although there were differences in initial achievement, how much students learned from each exposure to the material was almost identical for all students (Koedinger et al., 2023). This suggests that whatever the initial achievement of our students, all students can achieve at high levels if they are given enough instructional input.

Takeaways

- Building a bridge between teaching and learning requires that educators rethink aptitude and teaching.
- Educators and students can use assessments to improve learning.
- Assessments can be used formatively or summatively.