

Why there is no such thing as a formative assessment

Dylan Wiliam, *Emeritus Professor of Educational Assessment - UCL Institute of Education, UK*

What, exactly, is formative assessment? For some, it is a low-stakes assessment, given to students as a way to prepare for a more important assignment. For example, students might be encouraged to submit an essay before the final deadline so that they have a chance to get some feedback that they can use to improve the essay before it is finalized. This use of the term formative really describes where, in the teaching and learning process the assessment occurs. In this view, a formative assessment is any assessment before “the big one.”

For others, formative assessments are useful for checking on the progress being made by students. The idea is that we collect evidence of student achievement before the end of the teaching block so that we can check whether students are making enough progress to reach the required or expected standard. This is obviously a good thing to do. It is disastrous if the first time we find out that students haven't reached the required standards is at the end of the block, so we do need to be checking

that students are learning and remembering what they have been taught—we might call this kind of assessment “early-warning summative” assessment.

However, doing this is often not easy. If we assess students' progress half-way through a unit, do we assess everything in the unit, or only what they have been taught so far? If we assess everything in the unit, then students will obviously not do very well, because they are being assessed on things they have not yet been taught, but on the positive side, then most students will find that their final assessment will be higher than the interim assessment. If on the other hand, we assess only what has been covered by that point, then the students' scores will reflect how well they have learned what they have been taught, but if the second part of the unit is more demanding, this may give a misleading impression of their final level of achievement. There are no easy answers here; just trade-offs. The important point here is that by being clear about the trade-offs, we can make choices that are more likely to represent the best interests of our students.

As well as identifying students who might need more support to keep up with the rest of the class, used intelligently, such assessments can also identify mismatches between what students need to learn and what is being taught. For example, if



we have a spreadsheet that has students' names down the side, and question numbers across the top, with each cell containing a student's score on that question (zero or one for right/wrong questions, or the actual score for partial credit questions), then obviously, by adding scores horizontally, we get each student's score on the test. However if we add the scores vertically, we find out how well each question was answered. Low scores might indicate a topic that has not yet been taught, but might also indicate an item that the students should have been able to answer but were not. It might mean that there was something wrong with the question or that the approach to teaching what that question is testing was inadequate. The important point here is that collecting data about what students are learning helps teachers make their teaching more responsive to students' needs.

However, just thinking about formative assessment in this way—finding out what is going wrong after it's gone wrong—misses out on aspects that could make a much greater impact on students' learning.

While finding out whether students have learned something is important, assessment can play an important role in helping students become more active in managing their own learning—what Richard Stiggins calls “student-involved classroom assessment.” If students understand what they are meant to be learning, and are able to monitor their own progress towards those goals, they are likely to be able to improve their own learning.

Assessment can also help teachers make better decisions while teaching. The goal of all teaching is to increase long-term capability. After all, if students cannot recall something a week after they have been taught something, then they haven't really learned it. But teachers need to make decisions constantly while teaching—they cannot wait two weeks to see if the learning has “stuck”—and if they have better evidence about what is happening in students' heads, they are likely to make better decisions about what to do next. The fact that students can do something at the end of today's lesson does not guarantee that they will be able to do it in two weeks' time, but if they cannot do something at the end of today's lesson, it is rather unlikely that they will be able to do it in two weeks' time. Better evidence leads to better decisions which in turn leads to better learning.

In thinking about the quality of evidence that teachers have for their decisions, there are two particularly important aspects—breadth and depth. If a teacher asks a class of young children

which of the following are living:

- A. *Bird*
- B. *Tree*
- C. *Cat*
- D. *Table*

and asks students to raise their hands to indicate they have a response, then if one child chooses A and a second child chooses C, then it is tempting to conclude that the class has understood what a living thing is. However, if the teacher gets responses from only two students, then the teacher has very little idea about what is happening in the heads of the other children in the classroom. This is why it is a good idea periodically (and by that I mean roughly every 20 to 30 minutes of whole group teaching) to get a response from every single child, by using an all-student response system such as mini-dry-erase boards, letter cards, or even finger-voting (1 for A, 2 for B and so on). Some technology companies recommend using electronic voting systems, which they claim are particularly useful because it makes it possible to record every single student's response. However, I am not sure that this is such a good idea. If we want to create classrooms where students feel OK about making mistakes, the last thing we should do is record every single one of them. That is why the technology can sometimes get in the way and the low-tech version is often better, and which is why I sometimes joke that mini-dry-erase boards are the most important development in educational technology since the slate!

As well as getting evidence from every student, it is also important to pay attention to the quality of the evidence we are getting, and this is where careful design of the questions we ask is so important. If every single child in a class chooses A and C to the question above, then we might be tempted to conclude that they know what a living thing is. Unfortunately, this is unlikely to be a valid conclusion, since we know that many young children have a fundamental misconception about living things, and that is that they move. A child with this misconception would answer the question above correctly. Using the question below instead would reveal the misconception:

Which of the following are living?

- A. *Grass*
- B. *Bus*
- C. *Tree*
- D. *Computer*

The quality of the decisions we make depends crucially on the quality of evidence that we collect,



and this depends, in turn, on the quality of the questions we ask. After all, if students with the wrong thinking and students with the right thinking give the same answer to a question, it's not a very useful question to ask. This is why it is important that teachers plan the questions they are going to ask in class—ideally with a colleague—as part of their lesson planning. As one teacher once said to me, “You can't think up good questions on your own. You will always be victim of your own way of thinking.”

From the foregoing, it will be clear that formative assessment can occur over a range of time-scales from once or twice a semester at one extreme to minute-to-minute and day-to-day at the other. But it is also important to note that what makes an assessment formative is not when it takes place, or even the kind of assessment that we use. It is what we do with the information.



For example, suppose I test a child on his multiplication facts from 1×1 to 10×10 by selecting twenty of the one hundred possible number facts at random. If the child answers ten of the twenty correctly, I can be reasonably sure that he knows approximately 50% of his number facts. That would be a summative conclusion, because I am using the evidence to make a statement about that child's current state of knowledge. However, if I notice that he seems to be having particular difficulty with the seven-times-table, that gives me, as his teacher, something to work with. I know what to do next. That is a formative conclusion. The important point here is that the same assessment, and even the same assessment information, can be used summatively or it can be used formatively. This means that “formative” and “summative” cannot be properties of an assessment, but rather of the uses that we

make of the information, or the function that the assessment serves. If we think of assessments as procedures for drawing conclusions—we give students things to do, look at what they do, and draw conclusions—then the words “formative” and “summative” are best thought of as descriptions of the conclusions that we draw. Sometimes we draw summative conclusions (this boy knows 50% of his multiplication facts) and sometimes we draw formative conclusions (I need to help this boy with his seven times tables). Now to be sure, some assessments serve a formative function better, and some serve a summative function better, but if we accept that the words formative and summative describe the conclusions we draw, it is clear that there cannot be such a thing as “a formative assessment”, just an assessment that yields evidence that is used formatively.

To make this concrete, consider the interim test discussed at the beginning of this article. If we give a student a test half way through the block, then whether it is formative or not depends on what we do with the evidence from the assessment. If we score the assessment, and use that score to contribute to the final grade for the semester, it is functioning summatively, but if we also give the student feedback about what needs to improve, then it is also functioning formatively. The problem, of course, is that the presence of the score can often prevent students from looking at the feedback on how to improve. They look at their own score...and then they look at a neighbour's score—summative drives out formative. Any assessment can be used both formatively and summatively, but usually one function interferes with the other, so it is generally best to decide at the outset about the purpose of the assessment—is this to help the learner improve, or tell them how good they are? It's very difficult to do both at the same time.

