Formative assessment in 2050: Possibilities and challenges

Dylan Wiliam
UCL Institute of Education

When I advocate for increased use of formative assessment in classrooms, there are three questions I am regularly asked. The first is how do I know that formative assessment is not just another passing fad in education; one that in a year or two will be replaced by something else that catches people's attention?

My answer is that, as David Ausubel pointed out over 50 years ago, good teaching starts from where our students are, rather than where we would like them to be, and because what our students do learn from their classroom experiences is often unpredictable, we need to find out where they are in their learning before we decide what to do next. Put simply, assessment is the bridge between teaching and learning. It is only through assessment that we can find out the effect our teaching activities have had on our students. As long as teachers are systematically investigating the relationship between what they did, and what their students learned, they will always be able to advance their practice, and right now, the available research evidence suggests that there is no more powerful focus for teachers' professional development than this.

The second question I am asked is, after over thirty years of working in this area, whether my view of formative assessment has changed, and it really hasn't. Thirty years ago, now, and thirty years in the future, my vision is for classrooms where teachers are constantly making adjustments to their teaching, based on evidence from their students that is both deep and broad. By deep, I mean asking questions that go to the heart of what our students are thinking. In secondary school science, for example, rather than asking whether one's weight would be the same on the moon, which most students answer correctly, it is far more useful to ask whether one's mass would be the same on the moon, because many students get this wrong.

With younger students, when introducing decimals, teachers often ask questions that students can answer correctly with a faulty or incomplete understanding, such as, "Which is bigger: 0.2 or 0.3?" A better question would be to ask which is the bigger of 0.3 and 0.25, because many young children ignore the decimal point, and believe that 0.25 is bigger than 0.3 on the grounds that 25 is bigger than 3.

As well as asking questions that are worth asking, teachers also need to get responses from all the students in the group, rather than only from those who constantly raise their hands. After all, finding out what confident, articulate students think is unlikely to be a good guide to what is happening in the heads of the other students. Better evidence leads to better decisions which leads to better learning.

The third question I am regularly asked is about the role of technology, and here, I should say up front that I am sceptical. Many teachers use electronic voting systems or "clickers" to record a response from every student, but it seems to me that these systems often offer little over more traditional methods, and also have a number of disadvantages.

If a teacher wants to make a decision about what to do next in a lesson, then a well-designed multiple-choice question to which students respond by "finger-voting" (1 for A, 2 for B and so on) seems to me to be just as effective as using clickers, and, because most students have two hands, the teacher can include questions with two correct answers—something that it would not be possible with the clickers. Having multiple correct answers minimizes the likelihood of students getting the correct answer by guessing and if one of the correct answers is less obvious than the other, then this provides a way of stretching and challenging higher achievers. I have used finger voting effectively with groups of over a thousand students, and it provides me with a quick indication of the thinking of the students.

For constructed response questions, then mini-whiteboards, are very effective—indeed, I sometimes joke that mini-whiteboards are the most important development in educational technology since the slate—the best teachers were doing this hundreds of years ago.

What I don't have with the ephemeral evidence produced by finger voting or mini-whiteboards is a record of the students' responses, but this seems to me to be a feature, not a bug. After all, if we want to create classrooms where our students feel comfortable making mistakes, the last thing we should do is record every single one of them. And since I am using the students' responses to make a real time judgement about what to do, the lack of a permanent record is not a problem. Indeed, any data that I did collect would probably be out of date within a few minutes, especially if I am teaching well.

There are, of course, many other ways in which technology might help teachers, but it seems to me that the potential for such tools, at least in the near-to-medium term, is limited, for a number of reasons.

First, we have to remember that our aim as teachers is for long-term changes in what our students can do—if our students can do what we have taught them at the end of the lesson, but not next week, then our teaching has not been particularly successful. This is where the distinction that psychologists make between performance and learning comes in. Performance is how well students complete a learning task, and learning is the long-term change that results. This is important because it is possible for students to complete an instructional task unaided, and yet not learn what the task was intended to teach them. We can have performance, without learning. Perhaps even more strangely, we can also have learning without performance—students can be learning something and yet give no evidence that anything has changed. As if this were not hard enough, work by Robert Bjork and others has shown that students often learn more when they encounter 'desirable difficulties' in completing an instructional task; as Daniel Willingham points out, "Memory is the residue of thought". Technology that focuses on student performance on a task is likely to be a poor guide to what will be effective in the long-term.

Second, there is now a great deal of evidence that teaching that appears to be successful in the short term may be less effective, or even counter-productive, in the long term. Machine learning typically requires vast amounts of data and short feedback loops, and in games like Chess and Go, these are easily acquired. However, in teaching, the feedback loops can be years long, and so machine learning is unlikely to help much.

Third, the kinds of models we need to help us understand student thinking and learning are much more complex and detailed than it is possible to create currently. For example, to build a model of student learning in secondary school science that would be useful to teachers would probably require information on over 400 different specific capabilities, and several pieces of data would be required to know how far a student had progressed in terms of each of these capabilities. The amount of data that would be needed to create useful models of students' thinking are well beyond what is likely to be available.

Fourth, there is an ethical issue about collecting data on student learning. As mentioned above, useful models would require collecting data on our students all the time, and although some researchers have explored such "stealth" assessment, I think it is very important that students know when they are being assessed. This was neatly illustrated by a conversation I had with a 12-year-old boy named Lester that I taught some years ago. In a lesson on probability, he said that the probability of a flipped coin landing head up was 50%. As he was captain of the school football team, I asked him what he called during the coin toss at the beginning of the game. He replied, "Tails." I asked him why, and

he said tails came up more often. A stealth assessment of Lester's conversation would conclude that he did not understand probability, but he understands the mathematics perfectly well; he just doesn't think it applies in real life.

So what's to be done?

Perhaps the most important thing policymakers can do is to be sceptical about the use of technology to improve learning. While technology will undoubtedly have a role to play in our efforts to improve education, the fact is, in the words of Larry Cuban, technology has been "oversold and underused" for over 50 years. Given the long history of grandiose claims and disappointing results, evidence of sustained impacts in typical classrooms is needed before even considering technological solutions to educational improvement.

It is also important to understand what problems educational technology seeks to solve. Too often student data is collected, analyzed, and presented to teachers in the hope that teachers might do something useful with the information without any clear idea how the information might improve teaching. Rather than data-driven decision-making, we need to shift to decision-driven data-collection, by starting with the decisions teachers already make, and then figuring out what data might help them make those decisions in a better way.

Finally, policymakers need to be aware that the very nature of teaching creates enormous difficulties for developing effective technological solutions, because what we learn in one context might be a poor guide about what to do in another, similar context. It may be possible, at some point in the future, to understand and automate the decisions of expert practitioners, but for the foreseeable future, the professional judgement of expert teachers is likely to be superior to any decisions that could be made by even the most intelligent of machines. That is why the adoption of technology should take second place to the development of teacher expertise, through sustained, practice-focused, professional development.

Co-funded by the Erasmus+ Programme of the European Union