What assessment can—and cannot—do.

Dylan Wiliam

This article was originally published in Swedish translation as “Bryggan mellan undervisning och lärande” (The bridge between teaching and learning) in the September 16, 2011 issue of *Pedagogiska Magasinet*, a Swedish education journal.

Assessment is a central process in education. If students learned what they were taught, we would never need to assess; we could instead just keep records of what we had taught. But as every teacher knows, many students do not learn what they are taught. Indeed, when we look at their work, we sometimes wonder if they were even present in the classroom. In fact, it is impossible to predict with any certainty what students will learn as the result of a particular sequence of classroom activities. And because we cannot teach well without finding out where our students are starting from, we have to assess. Even if all our students started out at the same point (a highly unlikely situation!), each of them will have reached different understandings of the material being studied within a very short period of time. That is why assessment is the bridge between teaching and learning—it is only through assessment that we can find out whether what has happened in the classroom has produced the learning we intended.

Of course, assessment is also used for other purposes in education, which makes the picture much more complicated. In all countries, assessments of the performance of individual students are used to determine which students are, and which students are not, qualified for subsequent phases of education, and also to decide which kinds of education students should receive. In many countries assessment is also used to hold teachers, schools, and municipalities accountable to parents, taxpayers, and other stakeholders. I imagine that it would be difficult to find anyone who would argue that those who provide education should not be required to give some sort of an account to those who pay for education, and those who benefit, but in some countries, using assessment for purposes of accountability has produced significant negative consequences. In a small number of countries, teachers can be fired, and schools can be closed where the students’ results of high-stakes tests are deemed to be inadequate.

In some ways, this desire to use assessment results achieved by individual students to create a high-stakes accountability system is understandable. There is now quite strong evidence that the presence of a high-stakes accountability system raises student achievement by the equivalent of as much as an extra two months’ learning each year[[1]](#endnote-1). However, in *every single* instance in which high-stakes accountability systems have been implemented, adverse unintended consequences have significantly reduced, and in many cases have completely negated, the positive benefits of such an assessment system.

There are many reasons for these unintended consequences, but two are particularly important. The first is that accountability systems are rarely fair to teachers and schools. In every single country, the scores that students get at school depend far more on their individual achievement before they went to that school, the influence of socio-economic factors, and the support given by parents and other family members. For example, in Sweden, only 6% of the variation in students’ science scores in PISA in 2006 was attributable to the school; the rest was attributable to factors over which the school had no control[[2]](#endnote-2). In none of the countries that participated in PISA that year was more than one-quarter of the variation in student achievement in science attributable to the school. Holding schools and teachers accountable for something over which they have no control seems contrary to natural justice, and this is why many teachers and other education professionals find the idea of high-stakes accountability testing so repugnant. However it is possible to design systems of ‘intelligent accountability’ that control for the factors over which schools and teachers have no influence, for example by taking into account prior achievement, the socio-economic status of the students, their ethnic background, and so on[[3]](#endnote-3). When this is done, the traditional ranking of schools in terms of their results is overturned; schools that appear to be getting good results are shown to be complacent schools whose core competence is attracting able students, and adding little, while others, with modest results, are shown to be making extraordinary progress with students from disadvantaged backgrounds.

The second main reason for the unintended consequences is that because assessment results can serve a number of functions, there is a tendency to use the same assessment results to serve a number of functions, ostensibly to save time and money, and to reduce the burden of testing on students. While this is a laudable aim, using the same assessment information to serve different functions brings these functions into conflict, and frequently the result is that the assessment system serves none of the functions well.

One interesting example of this came to light in 1991. Using one of the commercially available standardized tests in the United States, a municipality found that its fourth grade students (those aged between 10 and 11) were scoring well above the national norms for their age. In fact they were about three months ahead in their learning compared with fourth-graders in the rest of the country. The following year, the municipality used a different test of the same subject, and was surprised to discover that on this new test, their fourth grade students were three months *behind* other fourth grade students. Over the next few years, however, the scores for fourth grade students went up, and three years after the new test had first been used, the scores of the fourth graders were, again, three months ahead of the expected scores for their age. That same year, a team of researchers gave a sample of the fourth grade students in the municipality the test they had stopped using four years earlier. The scores were three months behind the expected standards for fourth grade students. In other words, the teachers had become better and better at teaching the material on which their students were being tested, but the performance of the students on things that were not tested got worse[[4]](#endnote-4).

Another example comes from England, where a national testing programme was introduced in 1988. The scores that students achieved on these tests have gone up, but scores on tests used to compare performance in different countries, such as PISA and TIMSS have gone down. These two examples illustrate a more general phenomenon that is known as Campbell’s Law: “The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor”[[5]](#endnote-5). In modern education systems, we need ways of reporting on student achievement but when we use the results achieved by individual students to make inferences about the quality of schools, the assessment system becomes distorted so that the assessment results become less useful indications of what individual students have achieved.

The major reason for these distortions is that in most countries, assessments cover only a limited—and predictable—sub-set of the whole area they are meant to cover. If we want to assess a student’s capability in their mother tongue, we should assess their ability to speak and listen, as well as their ability to read and write, but few countries do this because it is time-consuming and expensive. Similarly, being good at mathematics involves the ability to explain mathematical reasoning, to design proofs, and to undertake extended investigations of mathematical ideas, but again, few countries do this.

Of course one way to make sure these ‘hard-to-assess’ aspects of each subject are included is to make greater use of the assessments made by teachers as part of their normal teaching, but this too has proved problematic in many countries. In the United States, the most important single piece of information used by universities and colleges to select applicants is the grade-point average (GPA), which is obtained by converting the grades awarded to students on their routine coursework into numbers (A=4, B=3, C=2, D=1) and average them across a course, and across courses. This results in what I call a ‘banking’ model of assessment in which once a student has earned a grade for an assignment, they get to keep that grade even if they subsequently forget everything they knew about this topic. It thus encourages a shallow approach to learning, and teaching. Students know that they only have to remember the material for two or three weeks until they take the test on that material, and they can then forget it, so there is no incentive for the student to gain the deep understanding that is needed for long-term recall. For their part, teachers tend to set tests that are straightforward, and that never go beyond what they have taught; because the tests are predictable, students who are well motivated do well, so both teachers and students feel good.

Of course the alternative system, in which all assessment happens at the end of the course, and is conducted by an external agency, as is the case in France, Germany, Japan, and many other countries is no better. In these countries, teachers have no role to play in assessing their students, which does leave the teacher free to concentrate on helping each student get the best score possible, but because the assessments are limited, ‘teaching to the test’ is possible, and often encouraged, and even if the teachers try to take a broad view of the curriculum, the students will often resist, asking, “Will this be on the test?”

The challenge, therefore, is to create an assessment that is externally-referenced, distributed, and cumulative. The assessment system needs to be *externally-referenced*, so that the teacher can honestly say to the student, “These are not my standards.” When the authority for the standard of achievement that students have to reach does not come from the teacher, the teacher is free to be a coach, rather than judge and jury. When the assessment is *distributed* across the whole course, the negative effects of ‘teaching to the test’ are minimized, because the only way to maximize the students’ achievement is to teach all the students everything. When the assessment is *cumulative*, there is no incentive for students (and teachers) to adopt a shallow approach, because if material is forgotten, it has to be learned again, because it is going to be assessed again.

There is no single best way to achieve this ideal of an externally-referenced, distributed, and cumulative assessment system, because any assessment system has to take account of the culture in which it will be used. A country in which great trust is placed in the professionalism of teachers will be able to get political support for systems that would be unpalatable in countries where such trust is lacking. The important thing is that the assessment system, as far as possible, creates positive incentives for teachers to teach well, and for students to study well. Once this kind of assessment system is in place, it should fade into the background and be unnoticeable, and unremarkable, because it would be so well aligned to the rest of the system. It would also support teachers and learners in focusing their time on using assessment to improve what happens in classrooms.

The idea that assessment should be used to improve learning is not new, but recently, many research studies have shown that using assessment during teaching, rather than at the end of teaching—what is sometimes called ‘formative assessment’ or ‘assessment for learning’ has a bigger impact on how quickly students learn than almost anything else[[6]](#endnote-6).

Although the terms formative assessment and assessment for learning are defined slightly differently by different people, there is increasing agreement that assessment improves learning when it is used to support five key strategies in learning[[7]](#endnote-7):

* Clarifying, sharing, and understanding learning intentions and criteria for success
* Engineering classroom discussions, activities, and tasks that elicit evidence of student achievement
* Providing feedback that moves learning forward
* Activating students as learning resources for one another
* Activating students as owners of their own learning

Each of these five strategies has a considerable research basis; together they provide a structure for ensuring that students and teachers work together to harness the power of assessment to improve learning

However, there have been many misinterpretations of these research findings, which prevent widespread adoption of effective practices. Perhaps the most widespread misconception is that any assessment that is intended to help learning will, in fact, do so. Many schools think that collecting data on their students’ progress and putting it all into a spreadsheet will help learning. There is absolutely no evidence that this kind of monitoring has any impact on students’ learning—as my American friends say, “Weighing the pig doesn’t fatten it.” At the other extreme, another widespread misconception is the idea that because a school has adopted formative assessment, there is no need to provide students with any indication of where they are in their learning. To be sure, giving grades and scores too frequently will certainly slow down learning, but not giving students any indication of whether they are making progress is just as misguided. The important thing about feedback, however it is given, is that it should cause thinking. Once an emotional reaction occurs, the learning will certainly stop, but there are ways of telling students whether they are, indeed making progress that does not allow them to compare their current achievement with others, thus minimizing the extent to which the student reacts to the feedback by attending to their well-being rather than by using the feedback to improve.

As I said at the outset, assessment is the central process in teaching. Without assessment there is no interaction—the teacher might as well be speaking to a videocamera that is being relayed to students in a different city. Assessment has a role in informing key transitions in education, and from education to work, and can play a role in assuring society that the money it spends on education is being used wisely (actually, it almost always is). But the most important assessment happens minute-by-minute and day-by-day in every classroom, and that is where an investment of time, and resources will have the greatest impact on student learning.

**Notes**

1. Wiliam, D. (2010). Standardized testing and school accountability. *Educational Psychologist,* **45**(2), 107-122. [↑](#endnote-ref-1)
2. Programme for International Student Assessment. (2007). *PISA 2006: science competences for tomorrow's world.* (Vol. 1). Paris, France: Organisation for Economic Co-operation and Development. [↑](#endnote-ref-2)
3. Ray, A. (2006). *School value added measures in England: a paper for the OECD project on the development of value-added models in education systems*. London, UK: Department for Education and Skills. [↑](#endnote-ref-3)
4. Koretz, D. M., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991, April). *The effects of high-stakes testing: preliminary evidence about generalization across tests.* Paper presented at the Annual meetings of the American Educational Research Association and the National Council on Measurement in Education held at Chicago, IL. [↑](#endnote-ref-4)
5. Campbell, D. T. (1976). *Assessing the impact of planned social change* (Vol. 8). Hanover, NH: The Public Affairs Center, Dartmouth College. Page 49. [↑](#endnote-ref-5)
6. For a summary of this research, see Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation,* **37**(1), 2-14. [↑](#endnote-ref-6)
7. Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree. [↑](#endnote-ref-7)