

# EXCELLENCE IN ASSESSMENT

## Assessment for Learning

31 October 2006

A supplement to the Cambridge Assessment Network  
*Assessment for Learning Seminar*  
held on 15 September 2006 in Cambridge, UK.

Compiled by Rebekah Oldroyd, Cambridge Assessment Network

## Excellence in Assessment: Assessment for Learning

### Contents:

Assessment for Learning: why, what and how.....	2
Professor Dylan William Deputy Director, Institute of Education, University of London	
Assessment for Learning: the challenge for an examination board.....	17
Neil Jones Principal Research and Validation Coordinator, University of Cambridge ESOL	
Assessment for Learning: putting it into practice.....	22
Laranya Caslin Deputy Principal, Sawtry Community College	
From Principles to Practice: making Assessment for Learning work in the classroom.....	26
Sylvia Green Head of Research, Cambridge Assessment	

A supplement to the Cambridge Assessment Network  
*Assessment for Learning* Seminar  
held on 15 September 2006 in Cambridge, UK.

**Cambridge Assessment Network**  
Cambridge Assessment  
1 Hills Road, Cambridge CB1 2EU  
Phone: +44 1223 553846  
Fax: +44 1223 552830  
[www.cambridgeassessment.org.uk](http://www.cambridgeassessment.org.uk)

**Cambridge Assessment** is the brand name of the University of Cambridge Local Examinations Syndicate, a department of the University of Cambridge. Cambridge Assessment is a not-for-profit organisation.

## Assessment for Learning: why, what and how\*

Professor Dylan Wiliam  
Institute of Education, University of London

What I want to talk to you about in the next hour or so is the *why*, the *what* and the *how* of using assessment to improve learning. *Why* should we be doing this, *what* should we be doing, and *how* do we go about it?

Here's the overview of the presentation:

- why raising achievement is important
- why investing in teachers is the answer
- why assessment for learning should be the focus of that investment
- how we can put that into practice.

### ***Why raising achievement is important***

I don't know if your workplace has those motivational posters that I find so irritating. This one is slightly different. The caption actually reads "If a pretty poster and a cute saying are all it takes to motivate you, you probably have a very easy job: the kind robots will be doing soon". I think that's a very important message because, although employers complain about the fact that school leavers aren't as skilled as they want them to be, there is no doubt that schools are doing a much better job than they've ever done before. What's going wrong is that the jobs that people need to do these days require much higher levels of numeracy, literacy and critical thinking than the jobs that were available 50 years ago. 50 years ago, the average working man or woman required neither to read nor to write during the working day; the kinds of jobs that you could do without those high level skills don't exist anymore.

Raising achievement matters. It matters for individuals: the premium is debatable but there's no doubt that the higher the education level you have, the more you earn during your life, the longer you will live and the better your quality of life. For society, the benefits are also huge. There are lower criminal justice costs; by increasing the level of education, you reduce the amount of money spent on incarcerating people. It reduces the cost of healthcare because people look after themselves better, and it actually increases economic growth.

Inevitably, quantifying the increase in economic growth you get from investments in education involves some pretty heroic assumptions. But Rick Hanushek calculated that, if we could invest in education and raise student achievement by one standard deviation over 30 years, the extra growth in the economy, and the additional taxes paid by people just because they were so much richer, would make compulsory education completely free of charge. Our economy would be so much bigger that we wouldn't actually have to pay for education from reception up to age 18. It would be free because of all that extra money coming in. The developing countries realise this. The developed countries realise this. The problem is that, if we are going to focus on becoming a knowledge economy, there are people on our tails and there is no alternative but to keep on raising the levels of educational achievement.

We've made huge progress in Britain. When I went to university in the 1970s, only about 5-10% of the age cohort went to university. Now it's 35-40%, and we also have very good retention and completion rates, unlike the rest of Europe. But, if you look at Singapore, Taiwan and South Korea, they're catching

---

\* Edited transcript of a talk given at the Cambridge Assessment Network Assessment for Learning Seminar on 15 September 2006 at the Faculty of Education, University of Cambridge.

up fast. So if we're serious about being as rich in the future as we are now, we have to carry on educating more and more of our population to higher and higher standards. So, forget OFSTED, forget league tables; I'm focusing on a much longer-term perspective than that. Even if changes of government mean changes of policy, the need to increase student achievement will be with us forever and that's the focus of today's talk.

So where's the solution? Small schools? Big schools? In England, there appears to be some support for making schools bigger. In the United States it's small schools that are all the rage because the Gates Foundation will give you lots of money to make your high school smaller. American high schools are huge—often over 3000 students, which makes them pretty impersonal places. That's why the Gates Foundation will give school districts money to build smaller high schools. So, Trenton, the state capital of New Jersey, decided to take their 3000-student high school and break it up into six 500-student high schools—in the same building. And they're wondering why nothing's changed. If you ask the advocates of small schools, "Why do you think high schools will work better if they're smaller?" they answer that students will interact with a smaller number of teachers. But that doesn't make sense. Students still have a maths teacher, a science teacher, an English teacher, a history teacher and so on; they have the same number of teachers whatever the size of the school. However, if you have teachers staying with the same group of students for several years, you might get some change, but the point is that none of those structural changes that we can make will make any difference by themselves. And yet there is a frantic search for a silver bullet that will raise standards quickly. Curriculum reform and textbook replacement are particularly popular in America because you can be seen to be doing something. "What are you doing to raise standards?" "We're going to get in a new set of textbooks." "Oh that's all right then". Other popular initiatives are changes in school governance, such as vouchers, but the evaluation of these reform efforts shows that once you control the demographic factors there's very little difference between charter schools, voucher schools, opted-in schools or opted-out schools. Those big structural things, those big easy policy levers to pull, have almost no impact on student achievement. They may make some of the changes you need to make easier or harder, but by themselves they achieve almost nothing.

The other big solution that's touted is technology but, as Heinz Wolff once said, the future is further away than you think. I think that for the foreseeable future we will have groups of between 20 and 40 students, with a teacher, and most of the learning is going to be in classrooms that are the size of classrooms we have now, with some IT of course, but the quality of the learning is going to be dictated by what's going on in that classroom. That's the big idea of this talk—if you're serious about raising student achievement, you have to change what happens in the classroom.

### ***Why investing in teachers is the answer***

As we get better and better value-added datasets, we're beginning to discover that the variability at the classroom level is up to four times greater than at the school level. So really there's no such thing as a good school, but there *is* such a thing as a school full of good teachers. Another way of saying this is that in terms of value-added it doesn't make that much difference what school you go to, but it matters very much which teachers you get in that school. If you get one of the best teachers you will learn in six months what it takes an average teacher a whole year to teach you. If you get one of the worst teachers, the same learning will take you over two years. There is a four-fold difference between the best and the worst teachers in the speed of student learning created.

This is a relatively new finding because, until recently, we didn't have the right kinds of datasets. To begin with, we thought schools *did* make a difference because some schools got good results and some schools got bad results. But then it was realized that different schools serve very different communities. Schools serving affluent communities got good results, and schools serving less affluent communities got worse results so, for a while, some people argued that schooling did not make much of a difference,

and the difference in results that we observed was due to demographic factors. However, once it was possible to look at the value-added by schools—the difference between what students knew when they began at the school and what the same students knew by the time they left—it became clear that the real differences are not at school level but at classroom level.

So what is it that makes such a difference in the amount students learn in different classrooms? It's certainly not the ability-grouping strategy. The research on between-class ability grouping—'setting' as it's usually called in the UK—is very clear. If you set, you will lower average levels of achievement. Setting raises the achievement of the higher attaining students but lowers the achievement of the lower attaining students because the bottom sets are more difficult to teach and typically get the worst teachers. However, because the losses of the lower achievers tend to be greater than the gains for the highest achievers the net result is a slight lowering of the average level of achievement, but that doesn't matter, because all the parents who are active and vociferous have their children in the top sets so that's okay. Within-class grouping also makes little difference, because what really matters is not how students are grouped. It's what happens in the groups, and that depends crucially on the quality of the teacher.

So if you're serious about raising student achievement, as the economic arguments, I think, mean that you have to be, then you have to be serious about improving teacher quality. So how do we do this?

Improving teacher quality is basically a labour force issue with two solutions. One approach is to try to replace existing teachers with better ones (this is the approach that Ronald Reagan tried with the air traffic controllers in America some years ago – sack them all and start again). It doesn't work in education because there's no evidence that more pay brings in better teachers. There are good reasons to pay teachers more, but there is little evidence that paying teachers more brings in better teachers. There is also little evidence that better teachers are deterred by burdensome certification requirements. Teachers brought in through alternative routes are in general no better than the ones who are already in post. So, to sum up the argument so far, if you are serious about raising student achievement, you have to get serious about improving teacher quality and if you're serious about improving teacher quality, you have to improve the effectiveness of *existing* teachers—what my colleague Marnie Thompson calls the "love the one you're with" argument.

The evidence is that it can be done. There are many small-scale studies that show that when you give teachers time, and appropriate kinds of foci on which to reflect, you can make teachers more effective. And by more effective I always mean in terms of student results. I don't necessarily mean in terms of student test results on National Curriculum tests but when I talk about a more effective teacher I always mean one whose students learn more. We may argue about how we measure that learning but as far as I'm concerned, the only purpose of having schools is to educate students and therefore the only evaluation of teachers should be based on how much their students learn. That's hard to do in a manner that's sympathetic to the broader goals of education but I think it is the only metric that counts.

So, as I said, we know how to support teachers in becoming more effective. From the pioneering work of Lawrence Stenhouse, we've known how to do it at small scale for over thirty years. The challenge is how we effect change across an entire local authority or even across a country.

If you look at the state of educational research, the situation is not very encouraging. At the British Educational Research Association Conference, or at the American Educational Research Conference, you will find researchers talking about the work they've done with 10 or 12 teachers, perfecting a vision of the perfect classroom. And while we don't have complete agreement about what would make the ideal classroom, there is substantial consensus on this. We know what the good classroom looks like. What we don't know is how to get lots more of them. And that is the main purpose of my talk today.

In thinking about the issue of improvement at scale, the first point I want to make is that we need to be clear about what we want. Guy Claxton uses the notion of learning power, while John Bransford talks about preparation for future learning but, either way, the fundamental recognition is that in preparing our students for life and work in the 21st century, what they know when they leave school will be less

important than what they learn later. And the key realization here is that *teachers do not create learning*—and as soon as you say it, of course, it's obvious. But almost everybody in the system functions as if teachers create learning. Teachers do not create learning, *learners create learning*, teachers create the conditions in which students learn. It's very hard to hang on to that idea, particularly as we become more and more accountable for children's results. There was a very interesting study done a few years ago by some psychologists. They were teaching mathematical problem solving to teachers, and split them into two groups. Both groups received exactly the same training except that, right at the very end, one group got one extra sentence: "Remember, you are responsible for making sure that your students perform to high standards". The only difference between the two training groups was that one sentence. When the researchers then went and looked at these teachers teaching, they found that the teachers who had heard this extra sentence exhibited more controlling behaviours in the classroom, told the students what to do rather than having them engage in the process of solving problems, with the result that the students learnt less about problem solving. Now if one sentence can make that difference to what teachers do, imagine the pressure that teachers are now under daily to do better. It's the old joke about schools being places where children go to watch teachers work and it's becoming worse and worse because we're under more and more pressure to do more. The hard thing is to say you get more learning by getting the students to do more of the work. You can't do anybody else's learning for them. We believe that in our heads, but we don't believe it in our hearts because, when the pressure is on, we revert to telling. At the time, it seems the right thing to do but we know it isn't.

The two features of effective learning power environments that are missing in most classrooms, both in the United States and the United Kingdom—and indeed in most other countries—are: one, that they create student engagement—what Lee Shulman calls "pedagogies of engagement"—and two, that they are well regulated; they keep student learning "on track"—what I call "pedagogies of contingency". The crucial ideas are that we create classrooms that engage students and that we control and manage the learning much more effectively than we're doing at the moment.

### *Pedagogies of engagement*

Why pedagogies of engagement? Well, unlike the impression that you probably get from reading the popular press, there is no doubt amongst psychologists that intelligence is inherited. If you looked at the popular media's coverage of the nature/nurture debate about whether intelligence is inherited, you would probably conclude that nurture has won out, but it hasn't. Everyone who knows the research evidence knows that intelligence is inherited to some extent. But so what? You can't change people's genes so the question is, is it entirely inherited? And the answer is no—it's also partly environmental. It's like physical height. Do taller parents have taller children? Yes, on average, they do. Does the genetic makeup of the parents determine how tall the child is going to be? No, because environmental factors like the quality of nutrition, particularly in early life, have a huge impact on how tall the child ends up being. And intelligence is exactly the same. There is a genetic component and an environmental component and we can't change the genetic component so let's not worry about it. What we *can* do is change the environmental component. So let's maximise that and let's give students the best nutrition possible for growing smarter students. Now the really interesting evidence is that intelligence, even when measured by IQ test scores, is much more malleable than we thought. The average British child now has an IQ about 15 points higher than it was at the end of the Second World War. It's called the Flynn Effect, after James Flynn from Otago University in New Zealand who was the first to notice this. IQ tests have to keep on being re-normed because each new generation of children is smarter than the generation before. So, basically, a student who would have got into a grammar school on the basis of

their IQ test score at the end of the Second World War would be below average today. Nobody knows why this is—nobody knows whether it's television, or diet, or something else. I happen to think it's television, have you ever watched a re-run of "Dallas" or "Dixon of Dock Green"? It's just so slow! So, although there is no doubt that intelligence is inherited to some extent, there's no doubt that we can actually make children smarter.

But here's the really new finding. As I just said, environment creates intelligence. Basically, the richer the environment, the smarter the kids grow up to be. But *intelligence also creates environment*. That's the new finding. Children (and adults) choose for themselves environments that match their preferred level of cognitive functioning. As I said earlier, the really important thing is that you create challenging learning environments—"high nutrition" environments to make students smarter, but if we look at the typical classroom it is clear some students participate and some students don't. Some students are trying to answer every single question the teacher asks. They're thinking all the time, and those students are getting smarter and smarter and smarter. There are other students who are trying to avoid getting called on. Guy Claxton called them "radiator children"—the children who snuggle up to the radiator and become almost invisible in the classroom. Teachers rarely, if ever, call on them, and so those students are foregoing the opportunity to get smarter. If you are allowing students to choose whether to participate in your classroom, you are exacerbating the achievement gap. So, because environment creates intelligence and intelligence creates environment, what we have to do is to create classrooms which are inclusive, where the level of cognitive demand is high, and where participation is obligatory. So if you're not requiring every single student to engage in the questioning you're doing or in the discussion you're leading, you are actually making things worse. That's why I don't think you have an option. If you're serious about raising achievement, if you're serious about getting rid of achievement gaps, between for example working class students and middle class students, then you have to create classrooms where participation is compulsory. Many students will resist because thinking is hard. Henry Ford used to say that it's extraordinary the lengths people will go to to avoid doing it. Thinking is really hard and that's the challenge: to create classrooms where it's not optional.

I want to tell you about the work of an American psychologist called Csikszentmihalyi who invented the notion of "flow". One of the interesting things I find in the way people talk about students is the way they talk about motivation. In particular, a lot of people talk about motivation as if it's a kind of parameter in a child's brain, like a dial with a needle. Some students have a lot of it and some students haven't got very much motivation at all. That's the prevailing view of motivation, that it's a kind of setting in a child's brain. The really powerful contribution that Csikszentmihalyi made was to think of it the other way round. To think of motivation not as a cause of achievement, but as an outcome.

My wife and I fostered teenagers in Hackney for about 15 years. One of the young people who stayed with us was struggling at school. His head of year wanted to send him to a special school for the delicate. One of the things the school said was that he couldn't concentrate for more than five minutes at a time. But we knew this wasn't true. We had bought him some construction toys—Lego and Meccano—and one Saturday afternoon he sat down for three hours and built an enormously complicated structure. I never was sure what it was, but we realized that the problems he was having at school were not the result of lack of motivation, but a mismatch between his capabilities and what he was being asked to do. When the level of competence is high, and the level of challenge is low, you get boredom, and when the level of competence is low, and the level of challenge is high, you get alienation. But when the level of challenge is just at the limit of your competence, you get this feeling of flow. Mountain climbers talk about "going pink"—when the level of challenge is just at the level of your limits of competence your concentration becomes almost infinite and time flies. Those of you who have experience of computer programming will know the feeling of saying "I'm just going to do another five minutes" and then three hours later you're still there, because you've become completely lost in the activity. Dancers talk about becoming lost in the flow of the dance and time becomes irrelevant. That kind of flow activity happens when the level of challenge is just at the level of competence. When those things are mismatched you either get alienation, boredom or apathy. When we start thinking about

motivation as being an outcome rather than a cause of student achievement, we actually look at our classrooms in very different ways.

### *Pedagogies of contingency*

Why pedagogies of contingency? There have been several major reviews of the research in this area. Gary Natriello in 1987, Terry Crooks in 1988, Kluger and DeNisi in 1996, Paul Black and I in 1998 and

Jeffrey Nyquist in 2003. This field is probably better mapped than any other field in education in the terms of what the big findings are, and the consistency of the findings is extraordinary. Whether you look at 5 year olds or 25 year olds, whether you look in Britain or in Portugal, whether you look in mathematics or modern foreign languages, you find the same effects and they're substantial. The central idea is that we should use assessment to influence learning and that the teaching should be contingent on what students have learnt, so that while we're teaching we collect evidence about where the students are to make adjustments to our teaching to better meet our students' learning needs.

If we look at cost-effect comparisons, reducing class size by 30% produces an effect size of 0.1 of a standard deviation. That's like giving students an extra three or four months learning per year, and it costs about £20,000 per classroom. So, does reducing class size increase student achievement? Yes, but not by very much and at huge expense. So the interesting question is not whether reducing class size helps, but whether it's cost effective. What's the cost, and what's the benefit? Increasing teachers' content knowledge by one standard deviation—i.e., moving the average teacher's content knowledge from where he or she is now to where the top one-sixth are—will give about the same effect, but we don't know the cost because no one's figured out how to do that. What we *do* know is that when you invest in teachers using formative assessment, or assessment for learning as some people call it, you get between two and three times the effect of class size reduction at about one-tenth the cost. So, if you're serious about raising student achievement which, as I said earlier, we should be, you have to invest in teachers and classrooms, and the way to do that is in teacher professional development focused on assessment for learning. This is not because I have a particular interest in it, but because if you're serious about doing it at the smallest possible cost this is the way the research indicates you should do it.

So far the argument's been entirely economic. The class size issue is quite complex but the reduction in class size shows very clearly that reduction, even from 40 down to 20, has little effect except when you're talking about reception and Year 1 and even then you only really get reasonable effects when you get the class size down below 15. So there *is* a case for having classes of 15 in reception and Year 1. After that, you'd be much better off investing in teacher professional development and teacher time. So, for example, teachers in Britain are contracted for 1265 hours, and if you look at what they're actually teaching, you find that most teachers in secondary schools, for example, are in front of students around 1000 hours a year. In Japan, secondary school teachers are in front of students for 540 hours a year. Even in Japanese primary schools, teachers only teach for 630 hours a year. They have classes of 40 in order to have plenty of time to work with other colleagues to develop really good learning materials. Whether you choose to have classes of 20 and no free time, or have classes of 40 and have 50% non-contact time, it costs the same. Which you prefer is a political decision. That's why I think the teacher unions that push for class size as the big issue are wrong. If we really want to professionalise teachers we should just be arguing that teachers should not be teaching the whole time. It's relatively easy to get more money for reducing class size. It's very difficult to get more money for having teachers not teaching the whole time, because the average voter believes that unless teacher are in front of students they're not adding value and that's the battle we need to be having. Barristers don't get paid only for the amount of time they're in court. Most of their work is done preparing for that and I think we need to mount a similar kind of argument for education.

## ***Why assessment for learning should be the focus of investment***

So, what is assessment for learning? Many people have come up with different kinds of formulations, but I would argue that there are five key strategies that encompass the terrain of assessment for learning or formative assessment. And I would say that if you're not doing one of these five strategies you're not doing assessment for learning, and if you *are* doing assessment for learning, you're doing one of these five things. The five key strategies are:

- clarifying and understanding learning intentions and criteria for success
- engineering effective classroom discussions, questions and tasks that elicit evidence of learning
- providing feedback that moves learners forward
- activating students as instructional resources for each other, and
- activating students as owners of their own learning

The “big idea” that ties these together is that we use evidence of student learning to adapt teaching and learning, or instruction, to meet student needs.

Before I go on, I want to clarify what I mean by the word “instruction”. To British ears, it conjures images of didactic teaching, of drill and practice. However, in America, the word “instruction” is a technical term that connotes the combination of teaching and learning. Indeed, in many languages, it is not possible to distinguish between “teaching” and “learning”. In Welsh, for example the word “dysgy” (pronounced “dusky”) means both to teach and to learn. If you were to say “I am teaching” and “I am learning”, you would say the same thing twice. From this perspective, the whole idea of separating out the quality of teaching from the quality of learning, as Ofsted has tried to do, makes very little sense. What are we to make of a lesson where the quality of teaching is good, but the quality of learning is not? It's rather like saying that the operation was a complete success but the patient died. So, when I talk about “activating students as instructional resources for one another”, I mean activating students as people involved in helping each other learn.

Formative assessment, or assessment for learning, is the pedagogy of contingency I mentioned earlier—the idea that teaching is constantly adaptive. A pilot guides a plane or boat towards its destination by planning a route, taking constant readings, making careful adjustments in response to wind, currents and weather. I flew back from Dubai last week—just imagine what would have happened if the pilot navigated the way that most teachers assess. We'd have set off from Dubai going vaguely northwest, and after six hours the pilot would say “Okay, it's time to land”, set down at the nearest airport and ask “Is this Gatwick?”. Whether it is or not, they say “I'm sorry, you have to get off now because we're on another flight tomorrow”. We teach stuff, and at the end of the teaching we give students a test and that's when we work out whether they've learned something or not, but it's too late to do anything about it because the national strategy says we've got to move on to the next unit tomorrow. So, this notion of keeping learning on track is the idea that the teaching should be constantly responsive to the students, so that if you get to the end of a lesson without having adjusted your teaching to take into account student learning needs you're already behind the game. It's about making your teaching constantly contingent on the students' responses. It's planning a carefully chosen, and possibly differentiated, route ahead of time—in essence, building the track—and taking readings along the way.

As I said, people use the terms “assessment for learning” and “formative assessment” in various ways. Paul Black, myself and the rest of the team at King's College London would say that there's an important distinction between the two. The distinction between assessment of learning and assessment for learning is basically about the intention behind the assessment. So, if you're assessing in order to help you teach better, that's assessment for learning, and if you're assessing in order to grade students, to rank them or to give them a score on a test, then that's assessment *of* learning. But in classrooms I see

plenty of what I would call formative intention but very little formative action. Teachers often say to me that they collect information in order to take action to help students, but if you follow it through, you find that the data never get acted on and the teaching never changes direction. That's why we say that assessment for learning becomes formative assessment only when the evidence of student learning is actually used to adapt the teaching work to meet student learning needs. If you're not using the evidence to do something that you couldn't have done without the evidence, you're not doing formative assessment.

There are different timescales for formative assessment. First, there's long cycle formative assessment, across units or terms. For example, you might collect evidence that shows that some students can balance chemical equations and some can't, so before the exam you go over this with the students; you're using evidence about student achievement to adjust your teaching over a long cycle. There's also medium cycle formative assessment within and between teaching units—a cycle length of one to two weeks. For example, you might give students a test before the end of the topic in order to be able to use the information to go over the difficulties before you finish the topic. But the research shows that the kind of formative assessment that has the biggest impact on student learning is short cycle formative assessment. Basically, if you're not using information to make a difference to your teaching within a day or two then it's unlikely to make a difference to student achievement. It's the short cycle formative assessment that really matters, minute by minute, and day by day.

Now what I've talked about so far isn't exactly new. I would have told you something very similar five years ago. I haven't changed our thinking about what formative assessment is or what it looks like or what our definition of effective practice is very much at all. What has changed quite a lot in my thinking over the last few years is what needs to happen for this to be put into practice. One of the things becoming clear is that we need much more sophisticated models of teacher learning. We need to be clear about the content of teacher change—*what* we want teachers to change—and the process—*how* teachers change. Very few attempts at reform have got the balance between those two things right. Many people are advocating teacher communities now; just put teachers in a room together and let them chat. That's a model that's strong on process, but unfortunately there's no evidence that it has any impact on student performance whatsoever because what teachers chat about matters. There are some things you can chat about that will actually make a difference to teaching and some things you can chat about that won't make any difference at all. You have to get the balance right. The national numeracy and literacy strategies were strong on content; they were clear about what they wanted to change about teachers, but they didn't have a good model of how that was going to happen. We need to have a balance between content and process. On the content side, we need to have evidence that this is the right thing to do and ideas for teachers to get started, and on the process side, the research says very clearly that we need to give teachers a choice of how to go forward, need to give them flexibility, provide small steps, make them accountable for making changes and give them support in making those changes.

### ***Strategies and techniques for formative assessment***

In talking about implementing formative assessment, I want to emphasize the distinction I make between strategies and techniques. The five strategies I mentioned earlier—things like engineering effective classroom discussions—are always good things to be doing; what Americans call no-brainers. Is it a good idea to involve students in their own learning? No question. Absolutely, yes. If I were a principal or headteacher of a school I wouldn't allow teachers to negotiate on whether they're going to work on assessment for learning or not. It would be required because the research shows that it is the most powerful lever for improving student achievement. But I would allow each teacher to choose how to do it. The reason I would allow choice here is that how you implement a strategy depends on the students that you're working with, and the subject you're teaching, and the kind of school you work in.

We would say that if you're serious about changing teachers, you have to get them to look at assessment for learning. It's more important than anything else. It's more important than classroom management, because classroom management without anything to direct it is effectively blind. It appears to be the most powerful focus for enabling teachers to change.

So the five strategies give you a range of things you could work on—you have to choose at least one of those—but then how you work on it is up to you. The strategies define the territory of assessment for learning but teachers are responsible for the choice of techniques for implementing that in practice. The important thing is that teachers need to adapt any technique that anyone else might show them to make it work in their local context. This creates ownership and shares responsibility for learning with the teacher. The key requirements of these techniques are that they must be grounded in deep cognitive principles about learning, they must be relevant to teachers' practice, teachers must see them as feasible (if teachers don't think that what they're being asked to do is really feasible in a classroom situation they won't do it), and they must be acceptable within the wider context. That's important, for example, if you are considering comment-only marking. There is no doubt that comment only marking is a good thing to do. It's just very difficult politically to do it when parents expect marks and grades on pieces of work, so we have to take that acceptability argument seriously.

So in our design process, we started in our own research with cognitive and affective insights into learning (affective means to do with emotions—an important part of learning). We then worked out how the five strategies fit together under the heading of the "big idea". Finally, we built up a series of techniques that teachers could actually use in their classrooms. However, our implementation process worked the opposite way. Teachers generally take on these ideas by starting with a selection of techniques, and by putting those into their practice they integrate them into their practice. Through the use of the techniques, the teachers then gain the cognitive/affective insights. The traditional model of teacher professional development is to start with the cognitive or affective insights of the researchers and to try to get the teachers to understand them, and then for the teachers to work out how to put them into their practice. We find that that doesn't work – it's just too hard. Instead, what we've done is think about how this might look in a classroom and to get teachers started there, on the principle that it's generally easier to get people to act their way into a new way of thinking than it is to get people to think their way into a new way of acting. Let's look at some of these techniques in a little more detail.

*Eliciting evidence of achievement.* One of the ideas that we've been developing to help teachers improve the way in which they find out what their students have learnt is the idea of a big question, which teachers work on very carefully in advance. For example, after teaching the students global warming, a teacher might ask, "What can we do to preserve the ozone layer?" and offer the students five alternatives:

- A. Reduce the amount of carbon dioxide produced by cars and factories
- B. Reduce the greenhouse effect
- C. Stop cutting down the rainforests
- D. Limit the numbers of cars that can be used when the level of ozone is high
- E. Properly dispose of air-conditioners and fridges

The teacher might then ask her students to hold up one, two, three, four or five fingers according to whether they think the answer is A, B, C, D or E. That's a pedagogy of engagement—the teacher is requiring every student to engage in this process, to think about the question and give her some information; after all if a student has not responded, it is very obvious. Then, if every student has responded correctly, she moves on. If no-one gets it right, she might teach it again, preferably in a different way. But if there is a lot of variation in the students' answers, she can direct the students to talk about their answers with their neighbours. This is a pedagogy of contingency. Her actions depend on the learning that is evidenced by her questioning. One teacher I have worked with tends to use four different alternatives, and has labelled each corner of her classroom A, B, C and D. Where the answers

are distributed across all the possible responses, she asks the students to collect together with the other students who chose the same answer in the appropriate corner, and the students plan together how they are going to persuade the students in the other corners that they are wrong. This is a pedagogy of engagement because the students have to get involved, and it's a pedagogy of contingency because the teacher is doing something that she couldn't have done until she knew what it was that each student thought was the correct answer. This question, incidentally, also illustrates another important point, which is that the rules for this kind of activity are very different from other testing contexts. For example, it is entirely appropriate, in a low-stakes classroom context, to include "trick questions" as above. The only correct response to this question is E, because it is a question about the ozone layer, not global warming. And yet, the correct response looks like a "makeweight" that has been inserted because the question setter couldn't be bothered to think of a proper fifth alternative.

Another example of a probing question, this time in English, is the following:

Which of these is a good thesis statement?

- A. The typical TV show has nine violent incidents
- B. There is a lot of violence on TV
- C. The amount of violence on TV should be reduced
- D. Some programmes are more violent than others
- E. Violence is included in programmes to boost ratings
- F. Violence on TV is interesting
- G. I don't like the violence on TV
- H. The essay I am going to write is about violence on TV

When they learn about persuasive writing, many students are confused as to what a thesis statement is and, for example, choose H because they confuse a thesis statement with an introductory statement. However, the other responses are more difficult to eliminate. One could make a case that several of these could be thesis statements, but most English teachers agree that, within the specific genre of persuasive writing, C is the best thesis statement. Of course, with eight responses, it is difficult to rely on students holding up a number of fingers to indicate their response, so many teachers have equipped their classrooms with a set of eight (or more) cards on a ring, attached by a string to each desk. Whenever the teacher wants to get a response from every student, she just says, "Reach for your cards" and each student has to hold up a card.

*Providing feedback that moves learners forward.* A further practical technique—comment-only marking—elicited confusion from a maths teacher who asked us how he could do that in maths. We suggested that instead of telling students that they got 15 out of 20, the teacher could, instead, tell them that five of their answers were wrong, and that they should find them and fix them. The important feature of this feedback, like comment-only marking, is that it engages students, leaving them with something to do. This technique was subsequently adopted by English teachers when they provided feedback on students' final drafts of writing assignments. Rather than correcting spelling, punctuation and grammar, the teachers put a letter in the margin for each error in that line using a G for an error in grammar, an S for a spelling mistake, a P for a punctuation, and so on. For the stronger students, the teacher would simply put a dot in the margin for each error, and for the weaker student, the teacher might indicate where in the line the error was. The idea is that the feedback gives something to the learner to do so that the immediate reaction of the learner is that they have to think.

Another way of providing feedback that moves learners forward is the idea of a 'three-quarters of the way through a unit' test instead of an 'end of unit' test. There's a problem with that because if, as a student, you get 95% on your test and your neighbour gets 45%, and then someone tells her what to do and she gets the same score as you, it's considered unfair. You should get a higher mark because you got it right first time. But think about that in the context of an MOT test for your car. My car passes first time, while yours fails on its catalytic converter. The garage tells you what's wrong with your car—they

don't just say bring it back when it's better—and when you get the problem fixed you get the same certificate as me. Is that unfair too? It's an absurd example but it shows that we're too locked into thinking about the purpose of assessment, the sorting and ranking and grading of students, than actually giving the teacher information about whether the class is ready to move on. It doesn't mean you wait for everybody because otherwise you will still be at unit one at Christmas but it does give you the information about whether the class is ready to move on, and you make that decision.

*Sharing learning intentions.* Many teachers provide students with lists of “success criteria” but these are often opaque to students. That's why it is particularly helpful to give students examples of annotated student work to “flesh out” learning intentions and success criteria (suitably anonymized of course). It is also very valuable to provide students with opportunities to design their own tests. There was a very interesting study done in 1994 where different groups of students were preparing for exams in different ways. Some students revised the materials they'd been studying, some students practised on “mock” tests and one group of students was asked to make up test questions (with answers!) on what they'd been learning. This last group got the highest score on the test. So if all you care about is cramming students for tests then the best way to do that is get them to create their own test questions.

*Students as owners of their own learning.* In one classroom where I've been doing some observation, every student has a disc which is red on one side and green on the other. When the lesson starts the green face is showing. The teacher goes through an explanation of the topic and if a student doesn't understand what's going on, they just flip the disc over to red. As soon as one student flips the disc over to red the teacher picks on a student who's showing green and that student has to come out to the front of the classroom and answer the question that the student who's showing red wants to ask. This technique is interesting because it embodies both pedagogies of engagement and pedagogies of contingency. In that classroom there is nowhere to hide because you're either saying you understand or you're saying that you want some help so that students are required to think about whether they understand or not (what psychologists call *metacognition*). The strategy is activating students as owners of their own learning, but it's also allowing the teacher to be responsive to the students' needs. Another teacher who tried this technique found that the discs were too difficult to see, so she went to the Party Store and bought red and green paper cups which were easier for her to see. It is really important that these techniques can be modified and adapted but it is also important that they embody the five key strategies, because otherwise we have no evidence that they are likely to be effective.

*Activating students as instructional resources for one another.* One technique that facilitates students helping each other in their learning is the “pre-flight checklist”. Before a student can submit, say, a lab report in science, the teacher requires the student to get a peer to complete a pre-flight checklist, which includes items such as whether the diagram is in pencil and labelled, whether it includes a title, a margin etc. The student can't hand in the report for marking until he or she has had this pre-flight checklist completed by a peer, and the peer has to sign that the check is complete. Then, if there's anything that's been missed on the pre-flight checklist that should have been there, it's the student who did the pre-flight checklist that's in trouble, not the person who submitted it. In this way one can force students to take seriously providing support for each other. The interesting thing about this technique is that it involves at least two strategies. It involves activating students as instructional resources for one another, but the person who completes the pre-flight check also has to understand the success criteria, in order to complete the pre-flight check. Furthermore, once students internalize the success criteria when assessing another student's work, it also enables them to use the insights gained in their own work.

The reason for distinguishing between strategies and techniques is that the strategies are always a good idea, but the particular techniques used to embody these strategies need to be chosen carefully, taking into account the subject matter, the students, and the context. For example, one technique for increasing student engagement during classroom questioning is to have the students' names on lollipop sticks, so that after you have asked a question, you can pick a name at random. This works well with younger children, but may not work so well for older students. Having said that, I have seen it used very

effectively with A-level maths students. After all, when you ask an 18 year old a question in front of their peers, the first thing that goes through their mind is “why are you picking on me?” With the lollipop sticks, the answer is that it is random. It’s just their unlucky day. They have to deal with it and answer the question. The important thing in all this is that the choice of technique must be left to the individual teacher.

### ***Putting it into practice***

So if what needs to be done is so clear, why is this so hard to put into practice? The answer is that we’re only just beginning to understand how complex teaching is, and that’s why research hasn’t changed the practice of teaching very much. That’s largely because most researchers have misunderstood the nature of expertise in teaching. Aristotle identified three main intellectual virtues – he called them *episteme*, *techne*, and *phronesis*. *Episteme* was the knowledge of universal truths, such as the fact that the base angles of an isosceles triangle are equal. That’s true for all time, you don’t need to go and check that tomorrow; it’ll still be true. *Techne* was the ability to make things, so the ability to make a table is not like *episteme* because there’s no perfect table, but the ability to bring a table into being is a very different kind of skill. But what Aristotle regarded as the highest intellectual virtue was what he called *phronesis*. This is sometimes translated as prudence but I prefer “practical wisdom”. The example of *phronesis* that Aristotle gave was the leadership of the state. The point he was making was that you should always have some principles underlying what you do but you must always temper those principles in the light of context, so there cannot be a set of unyielding, forever true, rules for the leadership of the state (the same idea is today enshrined in the principle that “hard cases make bad law”).

Why is this relevant to teaching? In education, “what works?” is not the right question because everything works somewhere and nothing works everywhere, so what’s interesting, what’s important in education is: “Under what conditions does this work?”. That is why teaching is mainly a matter of *phronesis* not of *episteme*. Teaching is all about knowing the conditions under which a particular technique is likely to work. That is why I say that you can’t tell teachers what to do. Not because I want to be nice to teachers but because the nature of expertise in teaching is not the kind of expertise you can communicate by telling people.

In fact, it turns out that a large part, if not most, of the knowledge that exists within the typical organization cannot be communicated by telling. Rather, much of the knowledge that exists in organizations is tacit rather than explicit. One definition of tacit knowledge is the kind of knowledge that an organization doesn’t know it has until the people who have it leave. This distinction between tacit and explicit knowledge was used by two Japanese authors, Nonaka and Tagueuchi, in a book called *The knowledge creating company* to look at how knowledge was transferred within organizations. For example, if you were the headteacher of a school, and you wanted to tell staff about new procedures that the government had introduced about the reporting of child abuse, then herding all the teachers into an auditorium and telling them about the new rules would be a pretty effective and efficient way to do it. Your explicit knowledge would be communicated as explicit knowledge to someone else.

However, most of the knowledge that’s important in teaching is not of that kind – most of the knowledge that we have in teaching is implicit. It’s not something we can actually put into words and so we need to think differently about how knowledge gets transferred. For example, sometimes other people’s implicit knowledge gets picked up implicitly, through a process of socialization. One learns how things work, without ever being told, and sometimes without the words to describe it. Sometimes, implicit knowledge becomes explicit knowledge through a process of externalization. For example, the first time I had student teachers in my classroom, I realized that I had a very poor understanding of how I did what I did. I ended up telling them just to do what I did. However, as I began to think about how to support student teachers, I became more reflective about my own practice, and as a result, began to develop explicit language for what I did—a process of externalization. The final process in Nonaka and Tagueuchi’s

model is the opposite—internalization—by which explicit knowledge becomes implicit. A good example of this is when one has been told something, but only after an extended period of practice and reflection does one get the sudden insight: “Aha, *that’s* what she meant!”. One can have the words for something but not really know what the words meant until they become internalized.

The important thing about this model is that it helps us think about creating knowledge in schools and, in particular, it shows the poverty of just telling people the solution. Now I want to make it clear that I’m not against telling people what to do in certain situations; if we want to increase teachers’ subject knowledge the best way is to put them into a classroom and teach them stuff. That’s a very effective and efficient model. It just doesn’t change practice. Every teacher I have ever met knows the research on “wait time”: the fact that students learn more if teachers wait at least three seconds after asking a question before providing hints or moving on to another student. The problem is that they just don’t do this when they are teaching. So the problem is not putting more facts into people’s heads, the problem is helping people live those practices in their classrooms, and the biggest problem is the simple fact that every single one of you learnt most of what you know about teaching before you were 18 years old. You learned about what it is to be a teacher, and the “scripts” of school in the 13 years you spent there as a student. The time you spent on a teacher training course is never going to shift those models. That’s why changing practice is so hard. We are talking about changing habits, not acquiring facts. Teacher professional developers have more to learn from Weight Watchers than from traditional educational psychology. Weight Watchers really understands that it’s about changing habits, not about changing knowledge. It’s about changing what people do day in, day out. If you’re serious about changing teachers’ practice in the classroom you have to help them change habits. And the way to do that is through small, school-based, teacher learning communities—groups of teachers that meet together regularly to support each other in making changes in their classrooms.

To recap, the argument so far is this: if you’re serious about raising student achievement you have to improve teachers’ use of assessment for learning; if you’re serious about helping teachers implement assessment for learning in their own practice, you have to help them do that for themselves as you cannot tell teachers what to do; and the only way to do that at scale is through school-based teacher learning communities. The good news is that you do not need experts to come in and tell you what to do. What you need is for you, as groups of teachers, to hold yourselves accountable for making changes in your practice. Implementing assessment for learning requires changing teachers’ habits. Teachers know most of what I’ve talked about today already so the problem is not a lack of knowledge, it’s a lack of understanding of what it means to do assessment for learning in practice. That’s why telling teachers what to do doesn’t work and why experience alone is not enough. If it were, the teachers who actually get the best value-added would be the most experienced teachers, and we know that’s not true. Most teachers don’t improve after about the sixth or seventh year in practice. That would be a very negative finding but for the fact that most of the inservice support that teachers are getting is not focused on changing practice. A typical five day in-service programme for a school year might have a day on differentiation, a day on personalised learning, a day on this, a day on that, and it’s all missing the point, because the only thing that matters is giving teachers time to hold each other accountable for making small incremental changes to their practice. People need to reflect on their experience in systematic ways that build their accessible knowledge base to learn from mistakes, and that’s what teacher learning communities (TLCs) are for. TLCs contradict teacher isolation, reprofessionalise teaching by valuing teacher expertise, deprivatise teaching so that teachers can talk about it, and offer a steady support for struggling teachers. (Here’s an interesting aside: I used to run a PGCE course and sometimes students would need to do an extra period of teaching practice because they hadn’t reached the standard after the first period. I learned that if they needed to do a second teaching practice the best place to put them was in the toughest schools because everybody there was struggling so they’d get much more support. The best thing that ever happened to me in my career was that my first school in London was on the White City estate in West London where everybody came out of their classrooms at 3.30 pm and said “Bloody hell”. Even experienced teachers were having problems keeping the students in the classroom,

let alone engaged in their learning. That was so powerful for me because we talked about things we might do, rather than pretending that everything was going well.) TLCs provide a regular space, time and structure for reflecting on teaching and practice; they facilitate sharing of untapped expertise and they build the collective knowledge base of the school.

Some people call these groups professional learning communities, but I prefer the term teacher learning communities, because in my view, if you are not a teacher, you cannot be a full participant in the community. The only people who can be full participants are those who are struggling to make these changes in their classrooms. Advisory teachers, heads, and others think they remember what it was like in the classroom but they don't. They can be peripheral participants and provide support, but they'll never be full participants because they don't know what it's really like.

So the synergy here is the idea of assessment for learning as the content of teacher change and teacher learning communities as the process. The model that I think would be implementable at scale is to establish monthly workshops about these kinds of ideas, so at the end of each meeting each participant promises her or his colleagues about what they are going to try out during the coming month. At the next meeting, everyone (everyone!) comes back to report on how it went. Repeatedly, we've had teachers comment on how silly they felt initially writing down their promise for the group, but expressing surprise that it does actually work. It's like Weight Watchers—promising to try something out and then being held accountable at the end of that process. It's what makes you prioritise developing your practice over all the other things that everybody tells you are priorities in school as well. Genuine peer observations, working with people at the same level as you in the system, also work well, when the agenda for the peer observation is set by the person being observed, rather than the person doing the observation.

I think this is a model for lifelong teacher development. It's what the Japanese teachers do through "lesson study". It's not something that you do in order to get good and then stop doing. The great thing about teaching is that you never get any good at it; you never crack it. That's what makes it so frustrating, so challenging, and yet so rewarding. Why is assessment for learning such an important part of this picture? Assessment is the bridge between teaching and learning. Assessment is the only way you can find out whether what you've taught has been learned.

## **Summary**

The argument I have been making is this: first, raising achievement is important; second, raising achievement requires improving teacher quality; and third, improving teacher quality requires improving teacher professional development. To be effective, teacher professional development must address what teachers do in the classroom and how teachers change what they do in the classroom. So the idea of assessment for learning combined with teacher learning communities turns out to be a uniquely high point of high leverage. It seems to be the way in which people can actually make changes in their practice most quickly; it acts as something of a Trojan horse into wider issues of pedagogy, psychology and curriculum. I'm not saying the other issues are not important – they are, but going up against them head first is unproductive. If you're serious about getting started, this seems to be the easiest way to make small changes to your classroom. The other point I want to make is that small changes are better than big changes because when you make big changes things go wrong, classroom discipline falls apart and you go back to doing what you know how to do, so you have to make very small incremental changes and not change more than one or two things. I would say that for most teachers, changing two or three things in a year is about all you could manage because otherwise you'd dart from one thing to the other. What you need to do is develop these things and make them so much of your practice that you don't need to think about them any more. Then, you're ready to take something else on. Once you've got into the habit, for example, of never asking a question without waiting three seconds, then you can take something else on, but if you still have to think to make yourself wait three seconds then you're

not ready to forget about that yet. What we're finding is that as well as making teaching more fun, and revaluing teacher professional expertise, formative assessment turns out to have some of the biggest impacts on student achievement, so not only will you get better test scores, you'll actually enjoy your teaching more.

## Assessment for Learning: the challenge for an examination board

Neil Jones  
University of Cambridge ESOL

### ***Assessment and learning***

Assessment bodies are becoming more involved in formative assessment—that is, assessment which feeds forward into learning. Being partly contingent upon political developments, this may be seen as a possibly transient phenomenon. Certainly the *No Child Left Behind* (NCLB) legislation in the US has led to a boom in school testing, also characterised as “no psychometrician left behind” (Jones et al 2005). But there are other, systemic, reasons for the coming together of assessment and learning. Technology is an important factor, with exam providers and publishers alike moving to exploit ICT. Hence the trend for assessment bodies to merge with media conglomerates, and for new technology-based consortia to enter the learning/assessment arena. Perhaps traditional summative assessment is simply becoming less relevant to the pace of change in a modern information economy. Or, as Bennett (1998) suggests, it may be rendered completely redundant by a future generation of intelligent learning environments.

In the UK at least this trend can be seen as the continuing evolution of a historical mission to support learning, dating from the mid-nineteenth century when the universities of Oxford and Cambridge set up boards to provide syllabi and examinations to schools. But how can an examination board extend its traditional summative assessment role to a formative one?

I will first present summative assessment from a specifically Cambridge ESOL viewpoint, and then consider how this approach might be developed for formative purposes. Two case studies are presented. The first is Asset Languages, a current DfES-funded project to develop an alternative assessment framework to support the UK’s national languages strategy. The issue here is how teacher assessments can be made to work formatively within a proficiency framework. The second by contrast is a research project into how an “intelligent tutor” for languages could be animated by an explicit model of learning and progression. These projects share a concern with the differing degrees of *support* offered to learners: summative assessment offers less, formative assessment more. The dimension of support is important to addressing the problem of comparability between these two modes.

### ***What is formative assessment?***

The term is used in at least two different ways. In one view, the difference between formative and summative is simply whether testing precedes or follows learning, and how the information is used, rather than in the design of the tests themselves. This contrasts with the conception of *assessment for learning* put forward in the UK by the Assessment Reform Group (1999). This advocates a move away from standardised testing and externally-defined criteria, towards assessment that improves learning, flexibly integrated into day-to-day classroom practice.

The first view is a convenient one for test providers, because it suggests that their existing constructs and methods might be economically re-purposed. In the second view, there might be no place for them at all. But it is the first view we must question, the second we must be prepared to explore. Though the values and techniques of summative assessment cannot be transferred uncritically to formative assessment, assessment bodies’ expertise can, perhaps, be turned creatively to these new challenges.

### ***A language proficiency framework which supports learning***

Starting from the Certificate of Proficiency in English, offered by UCLES in 1913, Cambridge ESOL now administers English language exams to over a million and a half candidates a year in 130 countries. From a single exam, its product range has developed to constitute a comprehensive system. The history of this development, like that of the influential Common European Framework of Reference (CEFR) (Council of Europe 2001) to which it is closely linked, illustrates a changing conception of assessment: from accrediting proficiency at an advanced level to a “learning ladder”, guiding learners from beginner level upward.

Cambridge ESOL exams report results in relation to a carefully-validated scale of communicative language proficiency. Proficiency testing uses a *trait* model: it represents a learner’s ability as a single point on a continuum. This conception has its critics. Indeed, an *item banking* (i.e. a trait-based) model for British educational assessment was rejected in the 1970s on the grounds that it represented too narrow a conception of learning and progression (Lacey and Lawton 1981). However, proficiency testing works and is useful when it validly supports inference to some “real world” beyond the test. Learners who pass Cambridge ESOL’s Preliminary English Test (PET) exam, for example, which targets CEFR level B1, can be confidently expected to perform in the real world in line with CEFR descriptors for that level, e.g. “...exploit a wide range of simple language to deal with most situations likely to arise whilst travelling. ...” (Council of Europe 2001, p74).

The concept of proficiency, although simple, is not simplistic. Constructs of proficiency are designed to be relevant to particular learning situations and purposes. Ability can be profiled by skill, providing individualised descriptions of performance. Moreover, nothing about the concept of proficiency, or the measurement models used to implement it, implies a narrow or naïve conception of language ability as something homogeneous and uniform. Proficiency scales are more about what differentiates levels than on what they have in common.

However, with its focus on the useful outcomes of learning, proficiency testing seems better suited to summative than to formative assessment. It captures progression in a way which motivates learners and helps them set broad learning objectives. It provides a valuable framework for learning, but not detailed information to feed into the next learning steps.

### ***Models of formative assessment***

But what kind of detailed information does assessment for learning require? One approach is to atomize high-level attainment targets into a checklist of detailed statements of attainment. This appears to characterise UK National Curriculum testing, and the general approach to NCLB formative assessment in the US. It is not a sophisticated approach and has been criticized for leading to test-driven teaching.

A more interesting model for formative assessment is proposed by Pellegrino et al (2001), who identify three elements of an “assessment triangle”: cognition, observation and interpretation. They argue that advances in understanding cognition (how learners represent knowledge and develop competence) and in the technology of assessment (using ICT to observe process as well as outcomes, sophisticated statistical models to provide interpretation), create new opportunities to devise assessments that can impact directly and positively on learning.

Pellegrino et al’s examples are mostly from maths and science and thus tend to focus on the process of problem-solving. For example, the steps in tackling a subtraction sum may include particular incorrect or incomplete strategies. Detecting these would give a truer picture of the learner’s knowledge than whether they simply get the correct answer, and enable a teacher to target the problem.

Clearly such insights could be of practical help to teachers. Pellegrino et al also give examples of “intelligent tutors”—computer systems which use cognitive models to provide targeted instruction and feedback. This is a qualitative rather than a quantitative approach to formative assessment, in line with Mislavy’s (1992, p15) statement: ‘Contemporary conceptions of learning do not describe developing competence in terms of increasing trait values, but in terms of alternative constructs’. However, it must also capture the notion of progression (which it does, often by contrasting the problem-solving strategies of “experts” and “novices”), and practically, one suspects, it may not always improve on the trait model. For example, basic maths tasks which demand different operations will tend to vary systematically in difficulty and should thus fit a trait model of progression.

For languages a more serious problem with the “alternative constructs” model is that applied linguistics has not provided usefully strong theories of cognition or language development. Language proficiency can, it seems, be reasonably described through a traditional division into skills—reading, writing, listening, speaking, language knowledge—with learners progressing steadily if unevenly in each of these traits.

So what model of formative assessment might work for languages? A different approach to analysing learning events is in terms not of the learning objectives themselves but of the situation in which learning happens. What learners do in the classroom is supported, or *scaffolded*. In a constructivist view, a learner’s state of knowing or understanding is inseparable from the particular situation (Teasdale and Leung, 2000). As Teasdale and Leung point out, this complicates the comparison of classroom performance with performance on formally-administered tests.

### **Case one: Asset Languages**

Asset Languages is a multilingual assessment framework in which two Cambridge Assessment divisions—OCR and Cambridge ESOL—are engaged under contract to the DfES. It implements the *Languages Ladder*, a proficiency framework emerging from the UK’s national languages strategy (DfES 2002), and constitutes a voluntary accreditation system supporting lifelong learning. Tests are available for three age groups from primary to adult, at all proficiency levels and in a wide range of languages. The framework complements existing qualifications by being based strictly on language proficiency, so that a given level (or *stage*) means the same thing across languages, irrespective of learning effort involved. The languages ladder stages are taken as broadly equivalent to CEFR levels, and so the framework can be seen as an extension across languages of the proficiency scale underlying Cambridge ESOL’s exams in English.

The scheme has two assessment strands: external at each stage, and teacher assessment—light-touch more informal accreditation—at finer grades (three grades to a stage). The teacher assessment strand should enable teachers to adapt flexibly their own materials and classroom practices for assessment purposes, within an orientation provided by the proficiency framework.

The initial approach uses a pack of materials for each language and level, comprising test tasks to select and use according to certain rules: teachers may adapt some of the tasks, for example, to use already-taught vocabulary. This is practical as a starting point, but we would wish to develop it further as a community of practice grows up among users of the system. Ideally the role of Asset Languages would be less to produce materials for teacher assessment, and more to facilitate the development and sharing of materials by teachers—achieving the “revival and reinvigoration of the principles and practice associated with ... graded objectives” envisaged by the Languages Ladder’s original proponents (Nuffield Languages Programme, 2002). The exam board’s role would be to ensure the alignment of classroom assessments to the framework and thus to the standards of the external assessments. These standards, it is worth stressing, can only be set and validated by focusing on the “real world” language skills of learners. Thus validation of external and classroom assessments are logically inter-related.

### **Case two: An online learning environment**

The second project is research oriented, relating to possible developments of an online learning project in which Cambridge ESOL is currently engaged.

The goal is a system which offers learners a path through a body of material, adapted to their current ability in a way which favours language acquisition. In the proposed model two orthogonal dimensions—difficulty and support—create a space within which learners can engage with tasks. The right combination of difficulty and support will favour learning—a notion that seems to relate to the *zone of proximal development* in socialisation approaches to learning (Vygotsky 1978).

The difficulty dimension captures progression in terms of linguistic features, topic range, aspects of competence, etc. The support dimension would emulate within the learning environment certain features of the classroom: sequencing (e.g. presentation, drill, practice) type of task (recognition, ordering, constructing), manipulation of tasks (length, redundancy, speededness, exposure), use of prompts, hints, reference materials, and so on.

The goal is to use the rich response data captured by the system to construct the difficulty and support dimensions empirically. This would constitute the learning model which the system could then use formatively, to adapt support and level of challenge to the individual learner, report progress meaningfully in relation to learning objectives, and provide positive feedback targeted at the learner's level.

### **In conclusion**

The Assessment Reform Group (1999) identify the following as crucial to formative assessment:

1. The provision of effective feedback
2. The active involvement of pupils in their own learning
3. Adjusting teaching to take account of the results of assessment
4. A recognition of the influence of assessment on motivation and self-esteem
5. The need for pupils to be able to assess themselves and to understand how to improve.

The two case studies I have briefly described in this paper indicate how an exam board might apply and build on the concepts and measurement techniques of summative assessment to help achieve these goals.

### **References**

Assessment Reform Group (1999). *Assessment for learning: beyond the black box*. University of Cambridge School of Education. Available at: <http://www.qca.org.uk/7444.html> (last accessed 27.10.06)

Bennett, R.E. (1998). *Reinventing assessment: speculations on the future of large-scale educational testing*. New Jersey: Educational Testing Service Policy Information Center. Available at: <http://www.ets.org/Media/Research/pdf/PICREINVENT.pdf> (last accessed 27.10.06)

Council of Europe (2001). *Common European framework of reference for languages*. Cambridge: Cambridge University Press

DfES (2002). *Languages for all: languages for life*. Available from: <http://www.dfes.gov.uk/languagesstrategy/> (last accessed 27.10.06)

Jones, L, Gueorguieva, J and Bishop, S (2005). *Some perspectives on the recent expansion in the use of assessments for formative purposes*. NCME Newsletter Vol 13 No 3. Available at: <http://www.ncme.org/pubs/news0905.pdf> (last accessed 27.10.06)

- Lacey, C and Lawton, D (1981). *Issues in evaluation and accountability*. London: Methuen.
- Mislevy, R.J. (1992). *Linking educational assessment concepts, issues, methods, and prospects*. Princeton: ETS
- Nuffield Languages Programme (2002). *A learning ladder for languages: possibilities, risks and benefits*. Available at: [http://languages.nuffieldfoundation.org/filelibrary/pdf/learning\\_ladder.pdf](http://languages.nuffieldfoundation.org/filelibrary/pdf/learning_ladder.pdf) (last accessed 27.10.06)
- Pellegrino, J.W., Chodowsky, N, and Glaser, R (Eds.) (2001). *Knowing what students know: the science and design of educational assessment*. Washington, DEC: National Academy Press.
- Teasdale, A and Leung, C (2000). *Teacher assessment and psychometric theory: a case of paradigm crossing?* Language Testing 17 (2).
- Vygotsky, L.S. (1978). *Mind in society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press.

## Assessment for Learning: putting it into practice

**Laranya Caslin**  
**Sawtry Community College**

My current role at Sawtry Community College, which is a Leading Edge school in north Cambridgeshire, is as Deputy Principal responsible for Learning Innovation, with a particular focus on Assessment for Learning. This article will outline the journey I followed from 2003 until the present day and I hope to share my experiences in such a way that you can draw on the positive lessons I learned and avoid the pitfalls!

I had been a member of Sawtry Community College's Mathematics Department from September 2000, whilst holding a middle management responsibility for whole school ICT. During this time I had seen a team of strong teachers each working hard to drive up standards of attainment; yet in that time our 'headline statistics' were getting steadily worse and reached an all time low of 54% A\*-C in August 2003. Ofsted was looming for the next academic year and so the Principal transferred me to Head of Department to do something about it ... in a term and a half!

One of the first things I did was to invite an Ofsted inspector to come and give us a health check to provide a starting point for what I knew would need to be a significant overhaul. The time was ripe; the department were bitterly disappointed/embarrassed about the very mediocre GCSE results and, as competent professionals, they definitely did not want to be in one of the failing departments in the school.

The inspector hit the nail on the head when he congratulated the team on being very dedicated and hard working, so hard working in fact that we were working much, much harder than our students! This gave us food for thought and was the catalyst for what turned into a revolution.

One of our pet whinges as a department was the fact that we were teaching the same topics over and over and the students seemed to show little level of retention. Many of the students were attempting to 'learn by osmosis' and just letting the lessons wash over them and we felt that we cared much more about their education than they did. Time to address the balance—and not by caring less ourselves!

We decided to use a wide variety of Assessment for Learning strategies to get the students to take ownership of their education. We wanted them to sit up and take notice; to appreciate the fact that every lesson (including the ones in Year 8 they thought didn't matter) was a small piece of a big 5-year education plan at Sawtry. We wanted them to understand their learning journey, the factors that determine their rate of progress and to take notice of the significant milestones along the way. Our rationale was to create a system of gathering and sharing performance information in such a way that we:

- empower the student to take responsibility for their own learning
- motivate the student to take responsibility for their own learning
- enable the teacher to direct their energy in the most efficient way

Here's how we did it:

At the start of the year Assessment Histories are produced for every class. They enable staff to understand what has already been achieved. In reality, these are produced in July so that teachers can see a snapshot of the prior attainment of their new class. They contain:

- base data such as KS2 results (including English to help staff note key literacy issues), MIDYIS data, FFT data
- information about what set the student was in last year so the teacher can tell who might be feeling a bit nervous, disappointed, over-confident, disillusioned having moved up or down sets at the end of the last academic year

- information about how each student did in each of the attainment targets so that teachers can see who might be especially numerate, who might have poor coordination skills or spatial visualization issues and who might have strengths or weakness in solving abstract problems
- information about progress made so far from KS2 and also last academic year—interesting to see who made significant progress during Year 7 but tailed off in Year 8 etc

The teachers go over this grid of information with different colour highlighters and it helps them to gauge the level at which to pitch their initial lessons and to devise an early seating plan.

At the start of the academic year, every student is given their starting level which is taken from their 'end of year' level from the previous year. We use a decimal system to help students to track their progress in greater detail. There are arguments for and against using a decimal system; but most of the arguments against it get bogged down with the diagnostic meaning of a decimal level. Level 7.1 simply means that the student was only a little way past the percentage needed to achieve a level 7 in the exam.

Students also understand that you can pick up marks from any of the questions in an exam covering a variety of levels. If there is a level 7 concept students find particularly difficult to understand, then there may well be a level 8 concept that is easy for them to take on board. Therefore, the marks for the level 8 skill could make up for the marks they may have lost on level 7. A classic example of this might be a topic such as 'tree diagrams', which many students find easy to understand. This might make up for a more difficult subject, such as 'simultaneous equations' which seem to be unfathomable to many of the younger students.

My students call these accessible high level marks 'bonus' marks. They love bonus marks and I have personally seen some C/D borderline students get so enthusiastic about them they turn straight to the back of the exam paper. Here they will find the favoured tree diagram question, the dimension analysis question and section about the power zero and the simple laws of indices, so that they get their bonus marks in the bank, before they start the rest of the paper!

The key point with the use of decimals to report performance in internal tests through Years 7 to 11, is that we then analyse where the students gain and lose marks.

By looking at the 'Start of Year Level', students are set an 'End of Year' target. Targets are set based on differentiated rates of progress. These targets will challenge the more able, yet are accessible to the least able and this will be allocated according to the students' individual learning set. The top sets are expected to add one National Curriculum level per year, and the lowest sets are expected to add 0.3 of a National Curriculum level per year.

I spent a long time working this system out by tracking past performances at Year 6, 9 and 11, for students of all abilities. I then calculated averages for small samples of students, who had obtained each GCSE grade and got all the work out just nicely on a graph with straight lines and different gradients. Total waste of time!

The first year we did this, virtually all of the students exceeded their targets—some making more than double the amount of progress I had calculated for them. In fact, the bottom set in Year 9 (who after eight years of education had only achieved a class average of level 3.1), made a class average progress of 0.9 level in one year.

The key thing is that the students have a target, something to aim for—most take the attitude 'well, I'm going to beat that!'. Some students cross out the target we set them and set their own 'apparently ridiculous' target. One classic example of this was a quite disruptive student, who had arrived in Year 7 on level 4, and at the end of Year 8 had regressed to level 3.6. I discussed his target with him; as he was in set 6 of 8 this was calculated to be  $3.6 + 0.5 = 4.1$ . I told him that he was capable of much more

than that and in fact, should be aiming for level 5. He defiantly declared that he was going to achieve a level 6, whether I believed him or not ... and he did. So after two years of drifting, he had gone down nearly half a level and then, after one year of motivation and empowerment, he had made up the lost ground and then gone on to achieve nearly 2.5 levels of progress. Another effect was that he stopped being disruptive, so that the students in his class were then able to focus on their work better too.

Obviously the process of knowing where they are now, and where they should aim to be by the end of the year, does not just make it happen. It gives them a focus, but that focus will be lost by the end of the second week of September if that is all there is! Our scheme of work is broken down into units which are about three weeks long on average. Each unit is launched via the skills tracker, which is a student friendly breakdown of the National Curriculum. These are differentiated, for example, at KS3 there are three versions: level 3 to 6, level 4 to 7, level 5 to EP. The range of levels is deliberately broad, so that they cover the likely range of levels for a student, across the whole of KS3. This is to emphasise the 'big picture' and to highlight to students when they might be doing work that is of a high level. We have found that students of all abilities love doing extension work, especially when combined with the fact that their teacher believes they will manage it—it is a bit like a vote of confidence and that gives them the boost to tackle it with determination and more often than not they succeed. The learning objectives for each lesson come directly from the skills tracker and the learning objectives are always written up for, and copied down by, the students in each lesson. This helps to focus the students' attention on the piece of the big jigsaw each day, and it helps them verbalise what they are good at and what they need help with.

Units of work also begin with a quick skills check activity that enables the students and staff to accurately assess the prior knowledge. This is such a simple idea, but it has been enormously powerful. The skills check is just a quick test (15 minutes) on core learning objectives from last year. This enables us to see how much they can confidently recall without any advance warning. Because we are working out what we need to teach them, the students are really honest about it—this stems partly from the fact that they haven't been taught anything on this topic this year so a poor score doesn't mean they are 'rubbish'. We use peer assessment and discussion to analyse what the class feels are the important concepts to revise, before moving on to new ground. Why is this so powerful? It is because the students set the agenda, they request the topics, they decide the starting point for their learning. The reality is that they usually request exactly what you intended to do anyway, but they are empowered. This means that next lesson when you say the learning objective is to 'revise angle properties of polygon' your response is 'oh yes, I need to do this' rather than 'boring, I did this last year'. The contrast to the apathy we were met with before we started this system is really potent, and the students are actually sitting up and taking notice of their very own, tailor-made education.

At the end of each half term, we conduct some formal assessments of the units that have been delivered. The tests reflect the national testing structure and we report results in decimal levels for each unit separately. All assessments are logged in an assessment book which stays in school and forms a record of their performance. The decimal level is then compared with the start of year level and the end of year target on a summary sheet. Parents are also informed of the results. By the nature of a mean average, all students will have their own relative strengths and relative weaknesses. We celebrate the strengths first and create the confidence to address the weaknesses.

Whilst we assess at the end of a unit, we do build in 'reflection time' in the scheme of work. This is really important if we want students to do something about the issues the assessment flags up. We have a feedback lesson for each test, whereby the students analyse their performance, question by question, against their skills tracker. Simple issues can be resolved immediately, sometimes as a whole class led by the teacher, sometimes in small groups with a similar problem, sometimes individually or by peer support from a more able student. Generic issues such as 'rounding answers correctly', 'showing working' etc can be set as a general target. In addition to the feedback lessons, we also have a window of opportunity, prior to the big end of year exam. The scheme of work has built into it 'mop-up/extension'

units, whereby students get their assessment books back out and have time to deal with more significant gaps in understanding. This again is organised by the teacher, using a mixture of whole-class, small groups work, peer support and individualised sessions.

The students sit an overall end of year examination, covering all the attainment targets, which assesses long-term retention and provides them with an end of year level. At the end of Year 7 and 8, we use the previous year's SAT exams, which also formed the Year 9 Mock. This provides interesting data, because all students in Year 7 to 9 have taken the same exam. Year 10 sit a set of slightly condensed GCSE exam papers.

You can appreciate that to set all this up was a significant undertaking, but the whole team pitched in and we got it up and running and embedded before Ofsted arrived in March 2004. The inspection went really well and we received a 'very good' overall grading with many 'excellent' comments—the inspector said it was the most 'excellents' he had ever put in a report. Considering our past performance in terms of examination results, this was a real achievement. Specifically relating to Assessment for Learning he said that 'Use of assessment ... is excellent' and 'Excellent procedures help to pitch students' quality of learning significantly beyond the good teaching'. We received a higher grade for Learning than we did for Teaching, which I understand is very rare and, when you consider that in September 2003 we were working much harder than the students, I think that this is proof that the balance had been addressed.

Obviously, Ofsted is not the be-all and end-all, but the successes did not stop there. August 2004 saw a rightly proud team of dedicated and hard-working maths teachers with headline statistics of Level 5+ KS3 results rising from 79% to 89% (and then 93% in 2005) and Grade C+ rising from 54% to 65% (and then 70% in 2005).

Since 2005 I have now taken the project to a whole school level. All assessments are now drawn together via spreadsheets linked into a database and we can generate complete Learning Profiles for students as often as we want—we have opted for once a half term. Form tutors, Heads of Year and parents are more actively involved in helping students manage their education. We are not home and dry—one of the difficulties has been getting teachers to embed the philosophy in their daily classroom practice and a downside is that, unless the teachers are teaching the students how to analyse their performance, then the process of sharing the performance data can actually be counter-productive. The crux of the message is that assessment is not Assessment for Learning unless you and the student analyse the information and use it to inform the next steps in learning. You don't fatten a pig by weighing it!

## From Principles to Practice: making Assessment for Learning work in the classroom\*

Sylvia Green  
Cambridge Assessment

I would like to begin by exploring some of the key issues surrounding the formative assessment and personalised learning debates that are taking place in England. The political message, from the Secretary of State for Education, is that no school should adopt a 'one size fits all' approach to teaching (*Education*, 2006). The aim is to ensure that schools cater for every child's talents and needs. In a paper presented to the General Teaching Council in England in 2004, David Hopkins, Chief Advisor to the government, proposed that

The most powerful lever we can pull at the moment to achieve personalised learning is assessment for learning ... it is a powerful means of helping teachers to tailor their teaching to pupils to get best improvement, and to involve, motivate and help them to take the next steps in learning. (p.10)

Whilst recognising the value of formative assessment he went on to report that

Although significant gains have been made and there are examples of outstanding practice, Ofsted identifies assessment and its application to teaching and learning as comparatively weak areas. Too many schools lack adequate systems for tracking the progress of individual pupils. (p.11)

Alongside this agenda is the ongoing debate about the value of formative assessment, a concept that is not new and was introduced by Bloom, Hastings and Maddaus in 1971. They put forward the idea that teachers should engage in formative assessment activities after stages of teaching and that, rather than relying on summative outcomes, they should provide feedback to students to target teaching and learning. Crook (2001) identifies five points that summarise the key lessons from research about formative assessment.

Assessment that promotes learning:

- involves learning goals understood and shared by both teachers and students;
- helps students to understand and recognise the desired standards
- involves students in self-assessment
- provides feedback which helps students to recognise next steps and how to take them
- builds confidence that students can improve their work

It can be argued that formative assessment is more educationally valuable and helps to provide personalised support by providing valuable feedback for students and data for teachers. One of the visions of the *Unified E-learning Strategy*, published by the Department for Education and Science in England (DFES, 2003), was to promote personalised feedback to help learners to progress by using immediate information about strengths and weaknesses, enabling support and targeting to be more effectively focused. With developments in the use of technology for assessment it is possible to automate processes of marking, recording and reporting, thus allowing teachers to spend more quality time with their students. One imperative in considering how to make the vision of personalised learning a reality is to ensure that any new models of assessment should not increase teachers' workloads to unacceptable levels. Assessment for learning and the assessment burden are two of the significant factors identified by Newton *et al.* as driving assessment policy and practice in England (2004).

---

\* Extracts from a paper presented to the International Association for Educational Assessment Conference, Singapore, May 2006.

The profile of **assessment for learning** (formative assessment) has steadily grown ... and has resulted in a number of policy initiatives and publications. It is being strongly advocated by many of the teacher associations which would like to see far less emphasis upon assessment for accountability and more upon assessment for learning. (p.46)

The final key driver of policy and practice in recent years has been the need to reduce (what has become known as) the 'assessment burden' to a minimum. This includes the burden of assessment processes upon pupils (the amount of time taken up by assessment as opposed to teaching and learning, the collation of assessment evidence, the recording of assessment results etc.). (p.46)

'The Tomlinson Report', on 14–19 Curriculum and Qualifications Reform (2004), proposed a reduction in the assessment burden by *introducing teacher judgment as the dominant method of assessing main learning at intermediate level and below* (up to 14 years of age) (p.85). The government White Paper, *14-19 Education and Skills* (2005), did not accept these proposals for replacing existing external assessment with internal teacher assessment. However, there was agreement that more robust teacher assessment can enhance the professional judgment of teachers and contribute to better teaching and learning and that formative assessment is an essential part of effective teaching. There was a commitment to provide training and guidance for teaching staff to develop their assessment skills and provide them with materials to help them to accurately assess student performance. There was also a commitment to keep e-assessment under continuous review and to exploit its potential to improve the quality of assessment and to minimise the assessment burden. (pp.68–70).

These developments and the increasing debate surrounding this area have the potential to open up a range of possibilities for assessment practices and will increase the potential for teacher assessment processes. However, there are a number of factors to consider in this context. Any attempt to introduce new methods or systems will require a balance to be struck that will inspire trust through reliability, validity and transparency whilst being practicable for teachers and students.

In the outcomes of the first year of a study carried out by the Assessment Reform Group (ARG), *Assessment Systems for the Future (ASF): the place of assessment by teachers* (2004), it is proposed that

Courses should provide more than information about what is involved in collecting and judging information for assessment. They should give trainees and teachers experience of generating, as well as using, criteria so that links are clearly made with learning goals. (p.11)

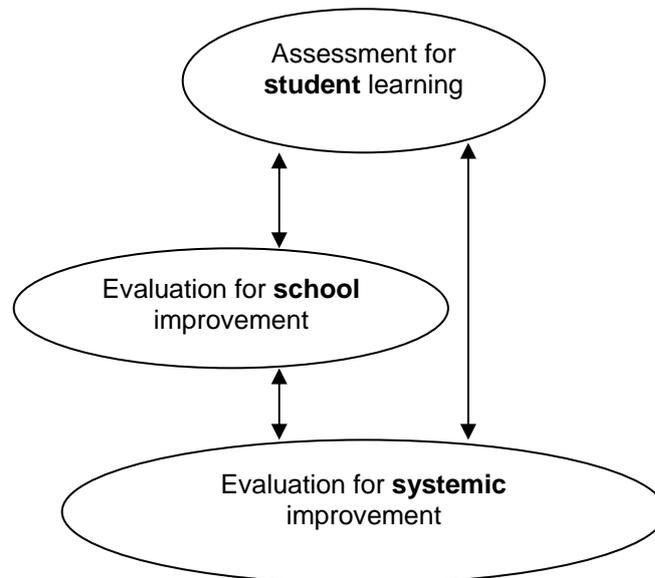
The Working Group Report also suggests that teachers' summative assessment should be explicitly linked to formative assessment, recognising the inter-relationship between assessment for the two purposes.

We can learn a great deal about different assessment strategies by looking at international practice and the analysis of the formative approach to assessment carried out by the OECD Centre for Educational Research and Innovation (CERI). In their publication, *Formative assessment: improving learning in secondary classrooms* (2005), the centre details classroom practice in eight countries since 2002 focusing on lower secondary education. They define formative assessment as *the frequent, interactive assessment of student understanding and progress to identify learning needs and shape teaching*.

They report very encouraging findings from their study but admit that there are still major barriers to wider practice. Such barriers include *resource and organisational implications related to practicality and tensions with the accountability demands of highly 'visible' summative tests of student performance*. The

study distinguishes between the levels of the education system at which strategies for improvement can be informed—classroom level, school level and policy level. (See Figure 1)

**Figure 1: Coordinating assessment and evaluation**



Source: CERI (2005). *Formative assessment: improving learning in secondary classrooms*. p.26.

The case study research by the CERI found that several OECD countries establish standards by providing detailed and shared criteria for students and teachers with internal systems used to track an individual's progression. There is international research evidence (p.47) that tracking a student's progress against objective criteria is more effective than making comparisons with other students' levels of performance. (Cameron and Pierce, 1994; Kluger and DeNisi, 1996; Heckhausen, 1989; and Rheinberg and Krug,1999).

Mischo and Rheinberg (1995) and Koller (2001) found positive effects where teachers referred to student progress over time. They reported that

The establishment of learning goals and tracking of student progress toward those goals makes the learning process more transparent; students do not need to guess what they need to do to perform well. Teachers also help students to track their own progress and to build confidence. (p.48)

Tracking progress was found to be an important part of the process as was the use of feedback if formative assessment is to work effectively. Good feedback should be linked to explicit criteria so that the learning process becomes more transparent and so that the student knows what is expected. Feedback also benefit teachers in that they can focus on the individual student's needs and consider how to plan the next steps in teaching and learning. Feedback combined with adaptation of instruction was important in the schools involved in the CERI case studies. A review of empirical evidence on feedback by Kluger and DeNisi (1996) showed that positive effects occur if feedback is formulated and used as a guide to improvement (p.225). Sadler (1989) emphasised that learners must understand both the 'reference level', that is, the goal of their learning, and the 'actual level' of their understanding. The

CERI study identifies an important aim of the formative assessment process as enabling students to evaluate and revise their own work (p.65).

Other areas investigated were peer and self assessment. Sadler (1989) argues that self assessment is essential to learning because students can only achieve a learning goal if they understand that goal and can assess what they need to do to achieve it. He suggests that self-monitoring is central to the work of all professionals and that students' self-assessment skills should be promoted if they are to become professional learners. Teachers at several of the CERI case study schools noted that peer assessment is an area where students need careful coaching and help in how to carry it out effectively. Students can be very critical of each other and require not only appropriate knowledge but also skills and sensitivity in order to use the process to best effect. However, when carried out positively, it can enhance the development of self-assessment skills.

Teachers appreciated the benefits of formative assessment while many at the same time recognised the barriers that can exist and that it is very difficult to put such ideas into regular practice. They need methods for translating abstract ideas into concrete practice and since they are very busy, with many pressures on their time, they need exemplars and tools to help them to gather information as part of the teaching and learning process. Nevertheless, those who managed to include formative assessment practices in their classrooms identified a number of benefits including

- improvements in the quality of teaching
- stronger relationships with students and increased contact with parents
- different and better work products from students
- greater student engagement

One of the key factors driving policy and practice, as identified by Newton, is the role of information and computing technology and the development of a range of electronic assessments. The collection, analysis and use of assessment results have also been facilitated and there is great potential for further innovation to provide support for teachers and students.

Research and development by Cambridge Assessment and Harcourt Assessment have led to the design of an online assessment system to support teachers and to engage students in assessments of Mathematics, English and Science. The system

- supports and informs teaching and learning
- is delivered online and is regularly updated
- includes assessment items and stimulus materials to motivate and engage
- provides instant marking
- gives instant feedback on performance to the teacher, pupil and parent
- provides personalised guidance on the pupil's next steps
- enables adaptive assessment appropriate for each pupil

At the beginning of this development, research carried out by Johnson and Green (2004, 2004b) investigated how assessment on-screen affected student performance, strategies, perceptions and behaviours. The evidence from their empirical study suggested that there were no statistically significant differences between student performance on paper and on screen and that there were fewer omitted answers on screen. The investigation of affective responses and motivational factors found that students felt that computer-based questions were easier than those presented on paper. The observation data showed that students showed greater commitment to answering on screen and were less easily distracted suggesting that they were more motivated and engaged with on-screen activity.

This research and development project offers the opportunity to collect a wealth of data on student performance across many areas of the curriculum and we intend to maximise the use of the data to investigate issues related to progression in the curriculum, common misconceptions and how they can create barriers to learning. During the next phase of the project we plan to continue the development of interactive, investigative activities to assess skills and processes in the contexts of Using and Applying Mathematics and Scientific Enquiry.

Feedback from teachers and students so far have been positive and the technology offers them support and saves time, particularly in recording and reporting. School managers also see the benefits as the system affords them the opportunity to gather data for target setting and organisation.

There is no doubt that personalised learning strategies and formative assessment practices can bring rewards in terms of student performance. However, they do place extra demands on teachers who will need training and support if they are to take on more responsibility in this context. Advances in technology will inevitably lead to more technological solutions in the classroom with support materials enabling more activity to be automated particularly in the areas of marking, recording and reporting.

It can be argued that formative assessment methods are more valid and educationally valuable but they also need to be trusted, reliable, dependable and practicable. A great deal of work is being done in the UK and internationally to promote formative assessment and we can learn a great deal from international studies about how best to support teachers and students in finding new strategies and new materials and in identifying the key areas that should be addressed in this context.

There is no doubt that new initiatives will need to be underpinned by thorough research so that we can make progress through evidence-based practice. With technology advancing rapidly and with the development of innovative assessment opportunities, the challenge will be to harness the technology and to make best use of all that it has to offer.

## **References**

- Assessment Reform Group (2004). *Aims and outcomes of the first year's work of the project*. ASF Working Paper 1. Available at: <http://www.assessment-reform-group.org> (last accessed 30.10.06)
- Bloom, B.S., Hastings, J.T., and Maddaus, G.F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill Book Co.
- Cameron, J and Pierce, D.P. (1994). Reinforcement, reward, and intrinsic motivation: a meta-analysis. *Review of Educational Research*, 64, 363-423.
- Centre for Educational Research and Innovation (2005). *Formative assessment: improving learning in secondary classrooms*. Paris: OECD Publishing.
- Crooks, T (2001). *The validity of formative assessments*. British Educational Research Association, University of Leeds. Available at: <http://www.leeds.ac.uk/educol/documents/00001862.htm> (last accessed 30.10.06)
- Department for Education and Skills (2003). *Towards a unified e-learning strategy*. Consultation document, July 2003. London: DfES.
- Department for Education and Skills (2005). *14-19 Education and Skills: White Paper*. February 2005. London: DfES.
- Heckhausen, H (1989). *Motivation und Handeln*. Berlin: Springer.

- Hopkins, D (2004). Assessment for personalised learning: the quiet revolution. In *Perspectives on pupil assessment*, a paper presented to the GTC conference, *New Relationships: Teaching, Learning and accountability*, London, 29 November 2004.
- Johnson, M and Green, S (2004). *Online assessment: the impact of mode on student performance*. Paper presented at the British Educational Research Association Annual Conference, University of Manchester, September 2004.
- Johnson, M and Green, S (2004b). *Online assessment: the impact of mode on students' strategies, perceptions and behaviours*. Paper presented at the British Educational Research Association Annual Conference, University of Manchester, September 2004.
- Kelly gets personal* (2006). In *Education*, No. 214, Friday 17<sup>th</sup> March.
- Kluger, A.N. and DeNisi, A (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.
- Koller, O (2001). Mathematical world views and achievement in advanced mathematics: findings from the TIMSS population III. *Studies in Educational Evaluation*, 27, 65-78.
- Mischo, C and Rheinberg, F (1995). Erzsziele von Lehrern und individuelle Bezugsnormen der Leistungsbewertung. *Zeitschrift für Pädagogische Psychologie*, 9, 139-151.
- Newton, P and QCA Statistics Team (2004). Assessment practice and developments in England with particular reference to the role of teachers in assessment for summative purposes. In Assessment Reform Group, *Assessment systems for the future: the place of assessment by teachers (ASF)*, Appendix F. Report of Seminar 2, March 29<sup>th</sup>-30<sup>th</sup> 2004, Cambridge. Available at: <http://www.assessment-reform-group.org> (last accessed 30.10.06)
- Rheinberg, F and Krug, S (1999). *Motivationsförderung im Schulalltag* (2. Auflg.) Göttingen, Germany: Hogrefe.
- Sadler, R (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Working Group on 14–19 Reform (2004). *14–19 Curriculum and Qualifications Reform: final report of the Working Group on 14–19 Reform*. London: DfES.

**Cambridge Assessment Network**

Cambridge Assessment  
1 Hills Road, Cambridge CB1 2EU  
Phone: +44 1223 553846  
Fax: +44 1223 552830  
[www.cambridgeassessment.org.uk](http://www.cambridgeassessment.org.uk)

**Cambridge Assessment** is the brand name of the University of Cambridge Local Examinations Syndicate, a department of the University of Cambridge. Cambridge Assessment is a not-for-profit organisation.