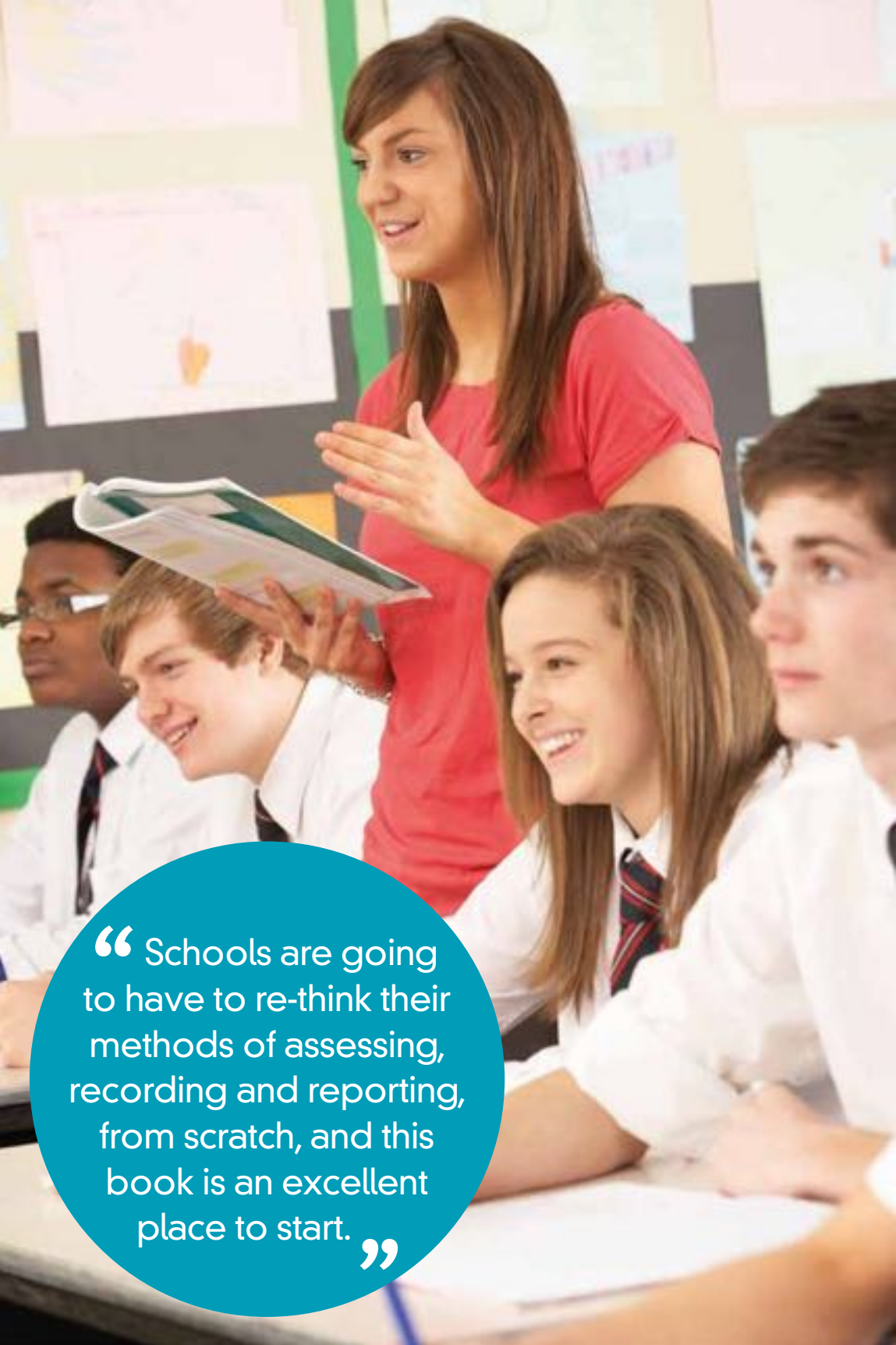

“ Schools are going to have to re-think their methods of assessing, recording and reporting, from scratch, and this book is an excellent place to start. ”

Foreword
by Dylan Wiliam

In July 1987, just as schools were breaking up for the summer holidays, the UK government announced its intention to introduce a national curriculum for all schools in England and Wales.[1] Over the summer, working parties were set up to propose what should be in the national curricula for mathematics and science, and, in parallel, the Secretary of State for Education and Science, Kenneth Baker, asked Professor Paul Black to chair the National Curriculum Task Group on Assessment and Testing.

The brief of the Task Group was to advise the Secretary of State for Education and Science, and the Secretary of State for Wales on 'a coherent system of assessment, including testing, to cover the whole period of compulsory schooling'.[2] The Group was specifically asked, in a letter from Baker, to ensure that its recommendations were practicable to implement and cost-effective, but should also:

> *take account of the very considerable amount of assessment which is already carried out as a normal part of teaching and learning in our schools, and will recognise that all forms of assessment affect the teaching and learning assessed.*

Baker went on to say that he was:

> *looking for arrangements which, by supplementing the normal assessments made by teachers in the classroom with simply-administered tests, will offer a clear picture of how pupils, individually and collectively, are faring at each of the age points. Such arrangements should help to promote good teaching.*

It seems that Baker had in mind a simple reporting scale whereby each student would get a grade from A to E at the ages of 7, 11, 14 and 16. Such a system of reporting would have the merit of simplicity, but the brief for the Task Group had specifically drawn attention to the fact that assessment affects teaching and learning. The problem with a simple A to E grading system was that a student who got an E at the age of 7 would probably also get an E at the ages of 11, 14, and 16.

At the time, I was working at King's College London, and had been closely involved with a project entitled Graded Assessment in Mathematics (GAIM). The project, directed by Professor Margaret Brown, was developing an assessment system that would meet the needs of all students, rather than just the 60% of students who would, in their final years of compulsory education, be taking courses leading to formal qualifications such as the General Certificate of Education (GCE, or O Level) or the Certificate of Secondary Education (CSE). In particular, we were looking at how to assess students who found learning mathematics difficult in a way that might be motivating rather than alienating.

I had come across the work of Carol Dweck in 1983, when I was teaching at North Westminster Community School, because she had contributed a chapter to a book edited by my headteacher, Michael Marland.[3] Dweck's distinction between 'fixed' and what she called at the time 'incremental' views of intelligence crystallized for me my concerns with many assessment models. If a student keeps on getting a C, then she is likely to come to think of herself as a C student. If the student keeps on making progress, the student is more likely to believe that intelligence is malleable—by working, you're getting smarter.

The idea of age-independent levels of achievement had been a strong feature of the Secondary Mathematics Individualised Learning Experiment (SMILE)[4] mathematics scheme that I had introduced at North Westminster Community School and so, when we were looking at reporting achievement in the GAIM project we tried to do the same thing—put all students on the same ladder, and ensure that all students experience progression.

Because we were focusing on the 13 to 16 age range, and in particular, low achievers within that age range, we felt that each student would need to achieve at least one level every year for the effects to be motivating. By use of archives of data from research projects such as Concepts in Secondary Mathematics and Science[5], and the reports of the Assessment of Performance Unit[6], Alice Onion—another researcher on the GAIM project—and I realized that to give every student in the 13-16 age range a reasonable chance of achieving one level every year, 13 to 15 levels of achievement would be needed.

Brown presented these ideas to the Task Group in the Autumn of 1987, and at one point, one of the group members asked how many additional levels would be needed to cover the primary age-range. Brown's estimate was that five additional levels would be needed. In other words, to give each student a reasonable chance of achieving one level a year, 20 levels would be needed.

While meaningfully identifying 20 levels of achievement might be possible in mathematics and science, no-one thought it would be possible in history or

English. The proposal of the Task Group was therefore to have a system of ten levels, designed so that the average student would achieve one level every two years. This made a lot of sense, because the intention was that student achievement would be reported only at the end of a key stage, so most students would experience progression in their reported level each time their achievement was reported.

Some years later, it was decided that the grades of the newly created General Certificate of Secondary Education (GCSE), which resulted from combining the Ordinary level of the GCE examinations with CSE, should continue to be used for the end of key stage 4, so the national curriculum assessment system was trimmed back to eight levels, covering students from 7 to 14.

This detail is important because it is often alleged that the ten (and later, eight) levels of the national curriculum assessment system were arbitrary, whereas the system was designed on the best research we had on student progression, and on the effects of grades and scores on student learning.

Over the following quarter century, the national curriculum was reviewed and updated many times, and on a number of occasions, the idea of age-independent levels of achievement was challenged. In his final report on the 1993 review of the national curriculum and its assessment, Sir Ron Dearing wrote: 'I am not convinced that the end of key stage scale provides a demonstrably better way to assess pupil achievement'[7] and the ten-level model survived.

Had schools continued to report national curriculum levels to parents at the end of each key stage (which is all they were ever required to do) then everything would have been fine. But over the succeeding years, schools started reporting levels at the end of every year, at the end of each term, and then, most bizarrely of all, schools started putting levels on individual pieces of work, displaying a staggering level of assessment illiteracy, since the levels were meant to be summaries of a student's achievement across an entire key stage. Even worse, inspectors from the Office for Standards in Education, Children's Services and Skills (Ofsted) would ask students what levels they were working at, and so, predictably, schools ensured that their students were able to respond with an appropriate number.

That is why, when, in 2010, I was appointed as a member of an 'Expert Panel' to advise Michael Gove, then Secretary of State for Education, on changes to the national curriculum and its assessment, I recommended that national curriculum levels should be abolished. I did so with no enthusiasm. I had always supported the idea of age-independent levels of achievement. I was convinced that such a system was more compatible with the idea of what Dweck now calls a 'growth mindset' and anything else would undermine efforts by schools

to persuade students that 'Smart is not something that you just are, smart is something that you can get.'[8] But what was happening in schools was so antithetical to good teaching, I thought whatever benefits national curriculum levels might bring were more than out-weighed by the negative consequences.

So, national curriculum levels have gone, and they will not be replaced. This has, predictably, been very disorientating for a lot of people, since many teachers and parents have known nothing else. But it is important to realize that the abolition of national curriculum levels represents an extraordinary opportunity for schools. It means that school inspectors can no longer ask how many students are making 'three levels of progress' or other equally fatuous questions. Schools will still need to ensure that they have good ways of finding out whether their students are making progress. When Ofsted inspectors ask, 'How do you know your students are making progress?' schools had better have a good answer to that question. But they are now free to choose ways of monitoring student progress that work best for them.

Most importantly, schools can develop assessment systems that take into account what we know about how learning takes place, and that is why this book by Daisy Christodoulou is so timely. As she makes clear, most school assessment systems rest on a profound fallacy—that the best way to monitor progress in learning is to judge progress by how far the student falls short of the level of performance that will be expected at the end of the learning.

The logic is attractive, but wrong. Over hundreds of years, we have learned that practicing scales when learning a musical instrument is helpful even if you will never actually play a scale in playing a piece of music. Sports coaches know the value of drills even though they seem remote from the kinds of skills that will be needed in competition. And in the same way, research is now demonstrating how to apply these lessons in academic learning.

For twenty years, I have been puzzling about the relationship between formative and summative functions of assessment. My initial instinct was that they could and should be integrated. After all, any assessment is just an attempt to determine what a student can do, and if the same assessments can serve both functions, then the time needed for assessment is reduced, leaving more time for teaching and learning. However in this wide-ranging and important book Christodoulou has, to my mind, convincingly demonstrated that while the formative and summative uses of assessments have to co-exist, they must also be kept apart. Record-keeping that details a student's progress towards test and examination success is unlikely to help achieve that success. Schools are going to have to re-think their methods of assessing, recording, and reporting, from scratch, and this book is an excellent place to start.

Notes

1 Department of Education and Science, & Welsh Office, 1987. *The National Curriculum 5-16: a consultation document*. London, UK: Department of Education and Science <http://www.educationengland.org.uk/documents/des/nc-consultation.html>
The above consultation resulted in the publication of the following report: National Curriculum Task Group on Assessment and Testing. 1988. *A report*. London, UK: Department of Education and Science. <http://www.educationengland.org.uk/documents/pdfs/1988-TGAT-report.pdf>

2 National Curriculum Task Group on Assessment and Testing. 1988. *A report: Appendix B* p.2 <http://www.kcl.ac.uk/sspp/departments/education/research/Research-Centres/crestem/Research/Current-Projects/assessment/tgatappen.pdf>

3 Licht, B. G., & Dweck, C. S., 1983. Sex differences in achievement orientations: Consequences for academic choices and attainments. In M. Marland (Ed.), *Sex differentiation and schooling* (pp. 72-97). London, UK: Hodder & Stoughton.

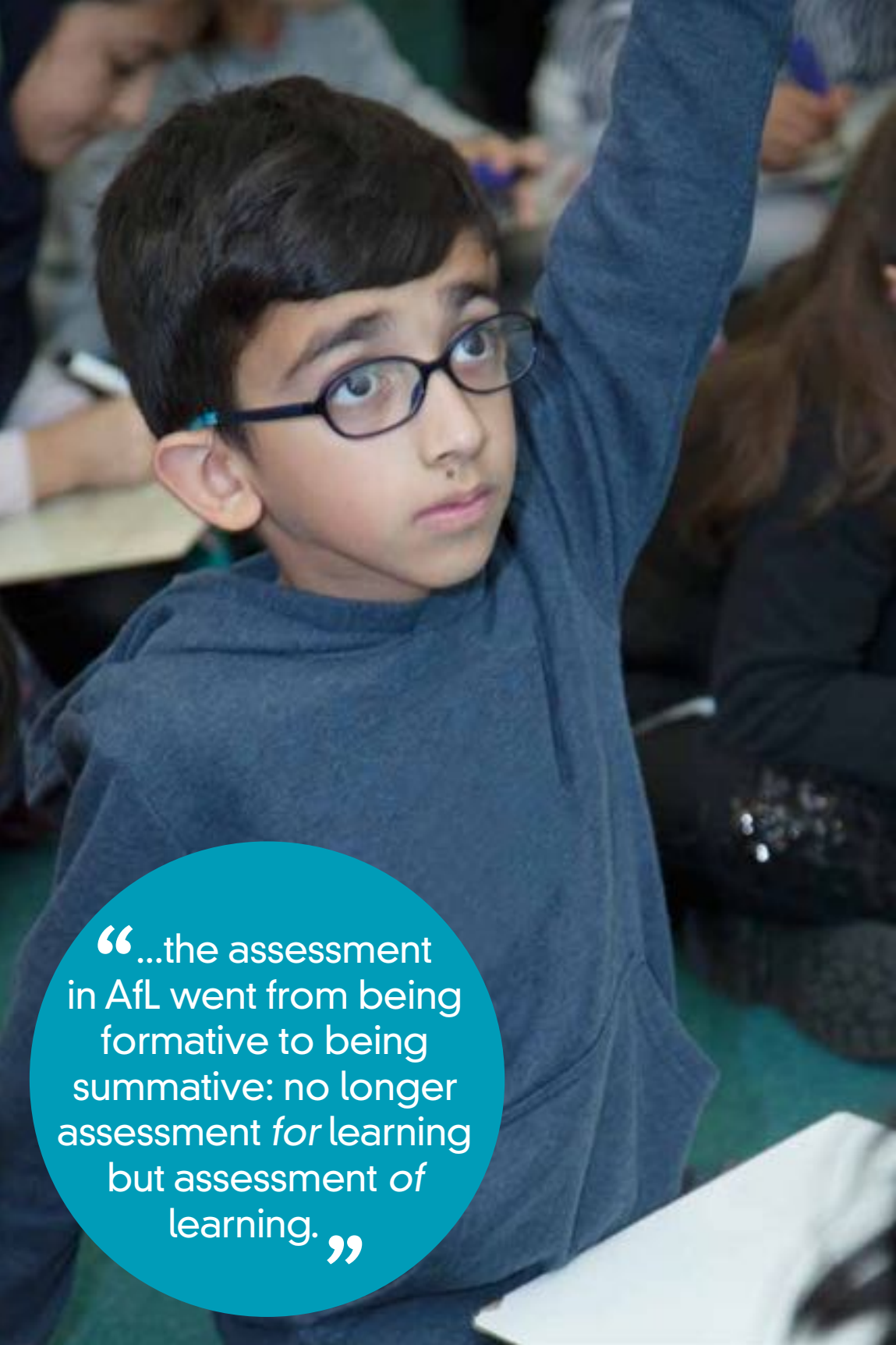
4 SMILE stood, originally, for the Secondary Mathematics Individualized Learning Experiment. Over the years this has been the subject of some revisionism, with claims that it stood for Secondary Mathematics Individualized Learning Experience, or even just Secondary Mathematics Individualized LEarning.

5 Hart, KM., 1980. *Secondary School Children's Understanding of Mathematics*. A Report of the Mathematics Component of the Concepts in Secondary Mathematics and Science Programme.
See also: Hart, K M. (Ed.), 1981. *Children's understanding of mathematics: 11-16*. London, UK: John Murray

6 For example: Foxman, D. D., Badger, M. E., Martini, R. M., & Mitchell, P., 1981. *Mathematical development: secondary survey report no 2*. London, UK: Her Majesty's Stationery Office (HMSO).
Foxman, D. D., Cresswell, M. J., & Badger, M. E., 1981. *Mathematical development: primary survey report no 2*. London, UK: HMSO.
Foxman, D. D., Cresswell, M. J., Ward, M., Badger, M. E., Tuson, J. A., & Bloomfield, B. A., 1980. *Mathematical development: primary survey report no 1*. London, UK: HMSO.
Foxman, D. D., Martini, R. M., & Mitchell, P., 1982. *Mathematical development: secondary survey report no 3*. London, UK: HMSO.
Foxman, D. D., Martini, R. M., Tuson, J. A., & Cresswell, M. J., 1980. *Mathematical development: secondary survey report no 1*. London, UK: HMSO.
Foxman, D. D., Ruddock, G. J., Badger, M. E., & Martini, R. M., 1982. *Mathematical development: primary survey report no 3*. London, UK: HMSO.

7 Dearing, R., 1994. *The National Curriculum and its assessment: Final report*. London, UK: School Curriculum and Assessment Authority p.70

8 Howard, J., 1991. *Getting smart: the social construction of intelligence*. Waltham, MA: Efficacy Institute p.7


“...the assessment in AfL went from being formative to being summative: no longer assessment *for* learning but assessment *of* learning.”

1 Why didn't Assessment for Learning transform our schools?

One of the most promising educational innovations of the last few decades was Assessment for Learning (AfL). AfL, or formative assessment, has been defined by Wiliam as when teachers 'use evidence of student learning to adapt teaching and learning, or instruction, to meet student needs.'[1] The concept of formative assessment was developed and popularised by Wiliam and Paul Black, two well-respected education professors. It was based on decades of solid research showing that giving feedback to pupils dramatically improved their progress.[2] Unusually, the idea was met with little opposition, and was in fact welcomed by both government and teachers. As Robert Coe, Professor of Education, argues, '[AfL] became the focus of national policy, widely endorsed by teachers and supported by extensive government training, following the publication of Black and Wiliam's (1998) *Inside the Black Box*… It is now a rare thing, in my experience, to meet any teacher in any school in England who would not claim to be doing Assessment for Learning.'[3] The 2013 Teaching and Learning International Survey (TALIS) confirms Coe's experience: in comparison to other countries, teachers in England give a lot of oral and written feedback to pupils.[4]

As Coe also goes on to argue, despite this propitious beginning, AfL has not had the kind of success you might expect:

> *…During the fifteen years of this intensive intervention to promote AfL, despite its near universal adoption and strong research evidence of substantial impact on attainment, there has been no (or at best limited) effect on learning outcomes nationally.*
>
> Coe, R., *Improving Education: A triumph of hope over experience,* p.10

Nor is Coe alone in claiming this. Wiliam and Black themselves have spoken of their disappointment in the way that the policy has been implemented: in 2012, Wiliam said that 'there are very few schools where all the principles of AfL, as I understand them, are being implemented effectively.'[5] Wiliam also noted that in many cases, whilst teachers had followed his advice in *Inside the Black Box* and replaced grades with comments, the comments they were providing were not necessarily that helpful – or that formative:

> *Typically, the feedback would focus on what was deficient about the work submitted, which the students were not able to resubmit, rather than on what to do to improve their future learning…I remember talking to a middle school student who was looking at the feedback his teacher had given him on a science assignment. The teacher had written, "You need to be more systematic in planning your scientific inquiries." I asked the student what that meant to him, and he said, "I don't know. If I knew how to be more systematic, I would have been more systematic the first time." This kind of feedback is accurate—it is describing what needs to happen—but it is not helpful because the learner does not know how to use the feedback to improve. It is rather like telling an unsuccessful comedian to be funnier—accurate, but not particularly helpful, advice.*
>
> Wiliam, D., *Embedded formative assessment,* p.120

Wiliam also said that, 'We have (Department for Education officials) saying: "We tried AfL and it didn't work." But that's because (they) didn't try the AfL that does work.'[5] How has this happened? Why did a policy with so much academic, government and grass-roots support end up being implemented so badly?

One possible explanation is that government support for the policy was, in fact, counter-productive. When government get their hands on anything involving the word 'assessment', they want it to be about high-stakes monitoring and tracking, not low-stakes diagnostics. That is, the involvement of government in AfL meant that the assessment in AfL went from being formative to being summative: no longer assessment *for* learning but assessment *of* learning. The difference may be just one preposition, but it is profound. When assessment is formative, the aim is to reveal pupils' weaknesses so the teacher can act on them. When assessment is summative, the aim is to give pupils a final grade, and so there can be pressure to try to conceal and gloss over misunderstandings. Indeed, formative assessment is so different from summative assessment that Wiliam has said that he wished he had called AfL 'responsive teaching', rather than using the word assessment.[6] He has also said that, 'The problem is that government told schools that it was all about monitoring pupils' progress; it wasn't about pupils becoming owners of their own learning.'[5] AfL is not just about teachers being responsive; it is also about pupils responding to information about their progress.

The pressures placed on assessment have almost certainly, therefore, played a part in the failure of AfL. Internationally, the Organisation for Ecomonic Co-operation and Development (OECD) has shown that all assessment systems struggle with the competing formative and summative functions of assessment.[7] In England, this problem is exacerbated by the pressures of a high-stakes accountability system. Schools are judged by how well their pupils perform on summative

terminal exams such as Key Stage 2 national tests (commonly known as SATs), General Certificate of Seconary Education tests (GCSEs) and General Certificate of Education Advanced Level tests (A levels). Not only that, but schools are also judged by the performance of their pupils in interim teacher assessments. When Ofsted inspect a school, they don't just look at the most recent national results. They also want to see the most recent teacher assessment data for pupils.[8] Therefore, there is clearly a great deal of pressure on these sets of data. Schools might want to set up internal assessment systems that aim to diagnose weakness, but the fact that the data in the system will be used by Ofsted to judge a school will make that much less likely.

Whilst there is undoubtedly some truth to this explanation, I do not think it accounts for the whole problem. Like many explanations which lay the blame at the door of Ofsted or government, it prompts another question: why did Ofsted and government distort AfL in this way? In this case, I think that the problems surrounding the implementation of AfL are the result of even more fundamental debates about the best methods of education.

In England, there is some consensus around the final aims of education. Literacy and numeracy are clearly vital skills which pupils need to be able to function in a modern economy and society. As well as these, developing skills such as critical thinking and problem solving are often agreed to be important aims of a modern education. Few people would be happy with an education which churned out pupils capable of reading basic texts and doing basic sums, but unable to think critically and creatively about problems they haven't seen before. Similarly, few would defend a system that ignored the basics of literacy and numeracy. The National Curriculum in England has been changed and revised since it was introduced almost thirty years ago, but its various versions stress the importance of the skills listed earlier.[9] However, whilst there may be some agreement on these aims, there is more controversy regarding the

best methods which will achieve them. We will explore these debates in more detail in the next chapter. For now, it is possible to summarise two broad approaches to developing such skills.

Teaching skills directly – the generic-skill method

One approach, which we will call a generic-skill method, is to teach a skill directly. If you want pupils to learn how to read, get them to read real books. If you want them to be good at solving maths problems, get them to solve maths problems. If you want them to think critically, set up activities and tasks that will give them the opportunity to think critically. In practice, such approaches might involve an element of project-based learning, where lessons are organised around skills such as problem solving, communication or critical thinking, rather than subject categories. So, for example, pupils might carry out a project where they work out the best place to site a new airport, or one where they design a leaflet to help guide people around a local museum. The idea is that if pupils work on solving problems which are more like the ones they might face in real life, this will help them to get better at solving such problems.

Teaching skills indirectly – the deliberate-practice method

An alternative approach, which we'll call the deliberate-practice method, argues that the best way to impart such skills is to teach them more indirectly. Whilst skills such as literacy, numeracy, problem solving and critical thinking are still the end point of education, this does not mean that pupils always need to be practising such skills in their final form. Instead, the role of the teacher, and indeed the various parts of the education system, should be to break down such skills into their component parts and to teach those instead. This means that lessons may look very different from the final skill they are hoping to instil. For

example, a lesson which aims to teach pupils reading may involve pupils learning letter-sound correspondences. A lesson with the ultimate aim of teaching pupils to solve maths problems may involve memorising times tables. The idea here is that the best way to develop skills may not always look like the skill itself.

The importance of debate about methods

These debates about educational methods are absolutely crucial to debates around formative assessment, because formative assessment is all about methods, whereas summative assessment is about aims. Or, to put it another way, different approaches to developing skill don't necessarily affect the assessment *of* learning. Because the outcome is less disputed than the method, the final assessment of learning won't look particularly different. We might agree, for example, that pupils should be able to write an essay about the causes of the First World War by the end of their time in school, perform a successful science experiment, talk intelligently about a character's motivation in *Macbeth*, or work out the standard deviation of a set of data.

Different interpretations of how we acquire skill really do affect assessment *for* learning. This is because these different interpretations are all about the *method* of acquiring skill. Assessment for learning is also all about the method, and the process, of acquiring skill. If we return to our earlier example, we may agree that pupils should be able to write an essay about the causes of the First World War by the end of their time in school, but we may differ on the process and the methods that will lead to them being able to write that essay. This will, therefore, fundamentally affect the assessment for learning that takes place as a part of this process.

If you subscribe to the generic-skill model, then very similar tasks can be used for assessment of learning and assessment for learning. The final assessment of learning should be the pattern for all teaching and all

formative assessment. If the final assessment is to write an essay about the causes of the First World War, then the formative assessment should also be to write an essay about the causes of the First World War, or perhaps to write a shorter version of the essay, or an essay about a related issue. This essay would then be marked and given feedback which would inform the pupil's next attempt at the essay. In this model, the assessment for learning tasks are very similar to the assessment of learning tasks. There are just more of them and they receive feedback. The result of this model is to do lots of tasks which have been designed to produce summative information but to add formative feedback to them.

On the other hand, if you believe that the methods that should be used to acquire skill are different from the skill itself (the deliberate-practice method), then assessment for learning looks completely different to assessment of learning. In this case, the terminal assessment is the end goal but the teacher or the curriculum designer must carefully break down that end goal into its constituent parts. So, if the aim is to get pupils to write an essay about the causes of the First World War, then formative assessments will consist of a range of different assessments which may look nothing like the final assessment. For example, formative assessments for this task may consist of short-answer questions that respond to a textbook article, multiple-choice questions about the causes of the war, activities that place key events in the build-up to the war in order, spelling tests on the key figures of the era, and narrative descriptions of key events. In this model, pupils may not even begin to write analytical prose until relatively late in the unit of work. In some units of work, they may never write any analytical prose, but the unit will still help to develop their skills of analysis by developing the skills and knowledge which underpin such analysis. With this model, most of the activities pupils do will not look like the final assessment, but the assumption is that these tasks will help pupils to do better on the final assessment.

One interesting implication of these different methods is that the generic-skill method is more likely to end up focussing narrowly on exam tasks because its model of skill acquisition suggests that practising a complex skill leads one to become better at it. The argument of this book is that assessment for learning became excessively focussed on exam tasks not just because of the pressures of accountability, but because the dominant theory of how we acquire skill suggested that was the best thing to do. The argument of this book is also that this dominant theory of skill acquisition is flawed. Not only has this model led to a narrow focus on exam tasks, it has also been ineffective at developing the skills that are its aim.

Because formative assessment is about methods, these debates about how pupils develop skill are crucial. It's therefore worthwhile considering the research around these different methods of skill development in more detail, which we will do in the next chapter.

Notes

1 Wiliam, D., 2009. *Assessment for learning: What, why and how*. London: Institute of Education, University of London

2 Black, P. and Wiliam, D., 1998. *Inside the black box: Raising standards through classroom assessment*. London: King's College

3 Coe, R, 2013. *Improving Education: A triumph of hope over experience*. Durham: Centre for Evaluation and Monitoring, Durham University, p.10

4 Micklewright, J., Jerrim, J., Vignoles, A., Jenkins, A., Allen, R., Ilie, S., Bellarbre, E., Barrera, F. and Hein, C., 2014. *Teachers in England's Secondary Schools: Evidence from TALIS 2013*. London: Institute of Education, University of London, p.154

5 Quoted in Stewart, W., 2012. Think you've implemented Assessment for Learning?. *Times Educational Supplement*, 13 July <https://www.tes.com/article.aspx?storycode=6261847> accessed 6 November 2016

6 Wiliam, D., 2013. Example of really big mistake: calling formative assessment formative assessment rather than something like "responsive teaching" [Twitter] 23 October <https://twitter.com/dylanwiliam/status/393045049337847808> accessed 6 November 2016

7 Organisation for Economic Co-operation and Development (OECD), 2013. *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*. Paris: OECD publishing

8 Ofsted, August 2015. *School inspection handbook: Handbook for inspecting schools in England under section 5 of the Education Act 2005*, p.12

9 For example, whilst the 2007 and 2013 versions of the National Curriculum had significant differences, in many ways the aims were quite similar: the 2007 English curriculum says that pupils should 'develop skills in speaking, listening, reading and writing that they will need to participate in society and employment… learn to express themselves creatively and imaginatively and to communicate with others confidently and effectively… learn to become enthusiastic and critical readers of stories, poetry and drama as well as non-fiction and media texts, gaining access to the pleasure and world of knowledge that reading offers'. The 2013 version says that pupils should learn 'to speak and write fluently so that they can communicate their ideas and emotions to others and through their reading and listening, others can communicate with them…. to develop their love of literature through widespread reading for enjoyment… write clearly, accurately and coherently, adapting their language and style in and for a range of contexts, purposes and audiences'. *The National Curriculum 2007*, p.61 <http://webarchive.nationalarchives.gov.uk/20100202100434/http://curriculum.qcda.gov.uk/uploads/QCA-07-3332-pEnglish3_tcm8-399.pdf> *National curriculum in England: English programmes of study*, 2013, p.2 <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/244215/SECONDARY_national_curriculum_-_English2.pdf> accessed 6 November 2016