

Free Schools for a Free Society

Reversing 30 years of failed reform

Tom Burkard and Colin McKenzie

A Report prepared for:

PARLIAMENT STREET

partnership in policy

About the authors:

Tom Burkard is a Visiting Professor of Education Policy at the University of Derby and is one of Britain's leading authorities on synthetic phonics. After teaching remedial literacy skills at Costessey High School in Norwich, he wrote the Sound Foundations decoding and spelling programmes. These are now used in well over 1,000 primary and secondary schools in the UK. He is the author of numerous articles and policy documents on education. He is a member of the NASUWT.

Colin McKenzie has achieved excellent results in Physics to GCSE and Chemistry to A Level for the last 12 years. He has trained teachers in basic literacy skills and provided resourced packages for scientific literacy and numeracy. As well as being Head of Chemistry and Key Stage 3 at Netherhall School in Cumbria, he played a leading role in improving science results at West Lakes Academy with a coordinated programme of testing and intervention. He now teaches at Pendle Vale College in Lancashire, where he is initiating objective tests of academic achievement in Science. He is a member of the NUT.

Never tell people how to do things. Just tell them what to do, and they will surprise you with their ingenuity. — General George Patton

Abstract:

In England, good teachers are leaving the profession at the same time that school rolls are increasing. Even good schools often have difficulty filling vacancies. Workload and poor discipline have made teaching a highly stressful occupation—and school improvement programmes can add substantially to workload. The problem is most acute in schools that serve the most disadvantaged children.

For the last 30 years, successive governments have made determined efforts to improve our schools, yet most employers and university admissions officers think that standards have fallen. This should not surprise us: solutions imposed from above are seldom implemented with any real enthusiasm. Excessive management is not the answer to a lack of inspired leadership: good teachers and future leaders do not thrive and grow under the shadow of Ofsted.

Unfortunately, our current means of assessing pupil progress are woefully inadequate. Teacher assessments—especially the time-consuming marking of workbooks—are not only subjective, but they are the major source of excessive workload. Most teachers readily admit (albeit in private) that much of their inputted data is false and that most of their marking is not for the benefit of their student, but for their observer. Ofsted inspections are an ever-present Sword of Damocles hanging over teachers and senior leaders, yet the evidence that inspections are followed by lasting improvements is weak to non-existent. This shouldn't surprise us: when lessons are frequently observed, pupils get the message: their teachers aren't real professionals—they're not trusted.

We need an **Ofsted-exempt School** programme where schools are actually free—one where teachers are only accountable for what their pupils have learned on objective tests of achievement, to be administered annually in core academic subjects. In secondary education, the basis for such testing already exists—many schools are already integrating routine tests of declarative and procedural knowledge into their lessons. These not only accurately measure the progress of all pupils towards good GCSE grades, but they utterly transform pupils' attitude to learning. Unlike the amorphous 'success criteria' in Assessment for Learning, their learning objectives are concrete and unambiguous; hence, even the least able pupils have hard proof that they are indeed 'making good progress'.

Initial reactions to this idea indicate that most good teachers would warmly welcome objective tests as an alternative to the Ofsted inspections, performance management reviews and subjective pupil assessment regime they are now expected to endure. Left with the freedom to develop their own ideas, England's best teachers could not only

vastly outstrip top-down measures but recruit and inspire a new generation of teachers; teaching could once again become a respected and attractive profession. Policy makers should welcome it too: at last we would have a valid and reliable measure of 'what works' in the classroom.

Introduction:

When James Callaghan voiced his concerns about educational standards in his 1976 Ruskin College speech, the DfE did not tell schools what to teach or how to teach—the *de jure* freedoms schools enjoy were still largely intact. The modern era of central control began when Kenneth Baker became Education Secretary in 1986. Since then, no expense has been spared—the Institute for Fiscal Studies has calculated that in real terms, education spending doubled between 1986 and 2010.¹

Judged by GCSE results, the return on this investment has been little short of spectacular. In 1987, only 26.4% of England's 16-year-olds achieved 5 A*-C grades. By 2012, this had increased to 81.1%. The National Literacy Strategy and the National Numeracy Strategy produced even more astounding gains: between 1995 and 2000, the number of 11-year-olds meeting the expected standard in English increased from 48% to 75%. Gains in Maths were even more impressive—42% 'passed' in 1995, and this increased to 72% in 2000.

However, if these figures represented real improvements, England's schools would have far outstripped the world's best. We'd be way ahead of Singapore—even the Finns and South Koreans would be flocking around our schools to discover our secrets. Sadly, Durham University's Centre for Evaluation and Monitoring (CEM) has demonstrated what any statistician already knew: in a system involving upwards of 7 million children and half a million teachers, improvement occurs slowly under the best of circumstances. On international tests, such as PISA, TIMSS and PIRLS, England's pupils have made few if any gains in the last 30 years.

Some independent studies paint an alarming picture of decline, especially in maths and science. One deserves special mention: a study by Shayler found that children's understanding of basic scientific concepts (such as displacement) has dropped sharply. When Year 7 pupils were given two identically sized blocks—one made of plasticine and

¹ https://www.ifs.org.uk/tools_and_resources/fiscal_facts/public_spending_survey/education Accessed 30/01/2017

the other of brass—they were asked if the metal block would displace more than, less than or the same amount of water as the plasticine block. In 1976, 54% of boys and 27% of girls gave the correct answer; in 2003 the figure was 17% for both. ²

Robert Coe was appointed chair of Durham’s CEM in June, 2013. In his inaugural lecture, he admitted that efforts to improve schools had met with little if any success. Professor Coe—and his predecessor at the CEM, Professor Peter Tymms—deserve full credit for throwing some light on a system which almost seems designed to defy scrutiny. Coe comes close to admitting that our current approach to education reform is not rational:

My argument so far has not been encouraging: things are not getting better, despite the best intentions and huge effort. Improving school systems is clearly very hard. Anyone who is convinced by the argument up to this point may be forgiven for thinking that the hope that we can do better in the future does not seem to be grounded in evidence. This part of my argument is not really evidence-based. In all honesty I do not know if we can do better; but I’m just not ready to give up yet. If we keep trying things that seem most plausibly likely to improve the outcomes we value, and - crucially - keep evaluating our progress, then eventually we must surely learn to make progress. ³

We should be asking why the last 30 years have seen unprecedented advances in science, technology and our standard of living—yet in education we have at best stood still. The current model—attempting to find which pedagogic approaches work best and inducing teachers to employ them—has been tested to destruction. All the while, we have largely ignored the most vital actors in this drama: the teachers who actually teach.

And therein lies the problem: there aren’t enough good teachers to go around. When failing schools can’t attract enough good teachers, school improvement programmes and CPD (Continuing Professional Development) merely serve to increase an already surreal workload. As a Guardian article argued,

Michael Barber's much-vaunted message that ‘the quality of an educational system cannot exceed the quality of its teachers’ is a powerful one that has led many school leaders and policymakers to throw more initiatives and training at teachers in order to make them better. The irony is that this approach leads to a loss of

2 Shayer, M, Ginsburg, D and Coe, R (2007) Thirty Years on—a large anti-Flynn effect? The Piagetian test Volume & Heaviness norms 1975-2003, *British Journal of Educational Psychology*, 77, 1, 25-41.

3 Coe, R C (2013) *Improving Education: A Triumph of Hope over Experience*, Durham, CEM, p xi

professionalism and autonomy, which simply fuels stress levels, reduces motivation and makes it less likely that teachers will have the capacity to improve themselves.⁴

We will not stem the exodus of teachers—let alone encourage talented new entrants to the profession—until we address the following issues:

1. **Top-down strategies don't work:** For the last 30 years, England's schools have been adventure playgrounds for ambitious education professionals. Many of their initiatives have worked well in tightly-controlled pilots, but when they are implemented on a larger scale they disappear without a trace. It is never very long until another exciting innovation comes along to displace the last one. Many of these ideas are too complex and too time-consuming to succeed, but the biggest problem is that teachers—that is, the ones who actually spend a good part of their day in front of a room full of squirming pupils—are not key players in the process of change. They are seldom consulted prior to the introduction of a new initiative or policy.

2. **Ofsted** is by far the biggest source of excessive workload. Despite attempts to ensure uniform inspection standards, individual inspectors still have their biases. In anticipation of their next inspection, schools have to prepare for a wide range of possible expectations. We argue that Ofsted is in fact the biggest *obstacle* to school improvement, and its remit should be limited to governance and finance. All of the problems listed below are to a greater or lesser degree a function of the climate of intimidation instilled by the inspection system:
 - i. **Proof of progress**—marking pupils' workbooks—the major source of excessive workload in most schools—does not lead to improvements in learning.
 - ii. **Performance Management** involves frequent observation of lessons. Observations are an unreliable means of assessing teachers' ability, and they undermine their authority. We cannot expect pupils to respect teachers when their managers openly demonstrate that they don't trust them.
 - iii. **School improvement programmes** tend to be shallow, repetitive and patronising. They often promote pedagogic concepts that lack a convincing

⁴ <https://www.theguardian.com/teacher-network/2013/aug/01/why-are-teachers-leaving-education>. Accessed 30/01/2017

research base, and it is rare for the programmes themselves to be independently evaluated. They do, however, add substantially to workload. Schools with the worst exam results are the most likely to engage CPD (Continuing Professional Development) providers.

- iv. **Discipline:** According to Ofsted, behaviour is at least 'satisfactory' in 99.7% of our schools. Yet at the same time, the Association of Teachers and Lecturers reported that 90% of their members had to deal with challenging behaviour such as shouting or swearing. Forty-three per cent said they had to deal with physical violence from a pupil in the last year. ⁵

3. Assessment: Despite all the disruption caused by inspections, they do not result in significant or lasting improvements in exam results. Nor do we have any objective evidence of what pupils are learning at any other point in their education—or even more importantly, what they *aren't* learning. Devising objective tests of knowledge and understanding is not technically difficult, but using them for accountability would be highly controversial. We believe that the most sensible way to introduce such a radical measure is to launch a new **Ofsted-exempt School** programme that would be open to existing schools as well as teacher-led proposer groups. In other areas of human activity, freedom has led to remarkable improvements—and these are almost always generated from the inside by people who have intimate knowledge of the activity concerned.

We now know that testing—which was an integral part of teaching and learning a generation or two ago—is by far the most reliable and efficient way to confirm and consolidate learning. The cognitive sciences have established beyond doubt that tests are much more effective than mind-mapping, re-reading, summarising or any other method of revision. ⁶ Most of our teachers deliver good lessons, but learning is quickly forgotten if it is not systematically reinforced. One of the more depressing aspects of teaching KS4 pupils (years 10 and 11) is finding out how little of what they learned in KS3 (years 7, 8 and 9) has been retained. **Ofsted-exempt Schools** will quickly appreciate the need for routine testing of previous material to ensure that that learning is consolidated in long-term memory. This is one of the major advantages of our proposal: it will create a powerful incentive for schools to focus on retention of what has been learnt.

5 ATL Media Office (21 Jan 2016) <https://www.atl.org.uk/latest/press-release/education-staff-facing-physical-violence-pupils-atl> Accessed 30/01/2017

6 Karpicke, J D & Blunt, J R (2011) Retrieval practice produces more learning than elaborative studying with concept mapping, *Science*, 331(6018), 772-775.

The failure of top-down reforms

Instructing the young is one of the most basic human instincts. We want to give them the knowledge and skills they need to survive and to inculcate them with our values and our culture. These will vary according to our race, religion, social class, political beliefs, education and experiences. It is foolish to think that ITT (Initial Teacher Training) will iron out all of these differences, and of course ITT courses vary considerably. Once an NQT (Newly Qualified Teacher) begins to teach, the nature of the school and the philosophy of its leaders will all make their mark.

With this in mind, it is easy to appreciate how difficult it is to implement any new idea. Inevitably, schools and teachers will implement it in the light of their own circumstances, experience and educational principles; very likely the result will look nothing at all like the originator intended. A study of federally funded initiatives in the United States found that there was little evidence that they were implemented consistently or faithfully.⁷ In a study of a Michigan reading initiative, Standerford found that “. . .changes attempted were rarely inconsistent with the teacher’s prior knowledge or beliefs about the best way to teach reading”.⁸

Coe admits that “. . .we do not know how to get large groups of teachers and schools to implement these interventions in ways that are faithful, effective and sustainable”.⁹ As an example, he cites Assessment for Learning, which was introduced in England by Black and Wiliam in their 1998 book, *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. And yet, the evidence presented above suggests that 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.¹⁰

7 McLaughlin, M W (1991) ‘The RAND Change Agent Study: Ten Years Later’ in Odden, A (Ed) *Education Policy Implementation*, SUNY Press, Albany, pp 143-55

8 Allington, R L (2001) ‘Does State and Federal Reading Policymaking Matter?’ in Loveless, T (Ed) *The Great Curriculum Debate: How should we teach reading and math?* Brookings Institution Press, Washington DC, p 269

9 Coe, op cit, p xi

10 Ibid, p x

Considering the extent to which AfL was promoted in England's schools, only a very brave teacher would openly admit that he or she wasn't 'doing it'. In fact, a report by the CfBT reveals that it has never really been implemented in most schools.¹¹ This should not surprise us: AfL is more an educational philosophy than a coherent set of classroom procedures. It demands that teachers radically change virtually everything they do, and at the same time the guidance they are given is vague and imprecise. The degree of 'personalisation' required imposes unrealistic burdens on teachers, especially in relation to 'feedback' – which, being easily observable in the form of marking pupils' workbooks, is one aspect of AfL which has survived and even prospered. For a secondary school teacher who has anything up to 200 workbooks to mark every two weeks, the workload implications of this requirement border on the surreal, especially when they are expected to provide detailed and thoughtful comments for each and every pupil as defined in this guidance:

Effective feedback should relate to the learning objective, pointing out success and improvement needs. It should offer clear guidance on how work can be improved, the next steps in learning and how pupils can take them.¹²

An essential feature of AfL is the active involvement of pupils in setting and achieving their own 'success criteria'. This also is very labour-intensive: among other things, it involves periodic interviews with each pupil to review their progress and agree on new learning objectives as a part of a personalised learning plan. This assumes a high level of pupil engagement. Considering the ATL's finding that 43% of education staff said they have had to deal with physical violence from a pupil in the last year – and of these, 77% have been pushed or shoved; 52% have been kicked; 50% have had an object such as furniture thrown at them, and 37% have been punched¹³ – it is not entirely surprising that some teachers might have difficulty persuading pupils to 'direct their learning' in the following manner:

The sharing of the learning intentions and the development of the success criteria is critical to the success of pupil assessment as this enables pupils to be objective about their own and others' efforts. The process of assessing another's work enables the pupil assessor to internalise the success criteria, giving rise to a deeper

11 Flórez, M T, & Sammons, P (2013) *Assessment for Learning: Effects and Impact*, CfBT Education Trust, Reading

12 declara, <https://declara.com/content/v5kK7nd5> Accessed 30/01/2017

13 ATL Media Office, op cit

level of understanding that can be transferred into his or her own work.¹⁴

Perhaps more to the point, 'learning intentions' in AfL relate more to generic skills than any body of knowledge. For the last generation, educators have been attempting to teach thinking and problem-solving skills that can transfer across the curriculum, and this is reflected in all of our assessment procedures. For instance, a pupil is more likely to be told that "you need to develop explanations of inferred meanings drawing on evidence across the text" than to be asked to analyse the ideas and information in the text itself. Not only does this fly in the face of evidence from the cognitive sciences, which have established that 'thinking skills' are domain-dependant,¹⁵ but it leaves the pupil with a task which is not only dry and difficult, but one where where pupils can never really be sure what is required or if they have succeeded. Ironically, this does nothing to make pupils 'independent learners'; rather, it renders them more dependent upon the over-stressed and over-worked teacher. Hence, it is hardly surprising that in reference to a 2008 Ofsted evaluation of the implementation of AfL in 43 schools, the 2013 report by the CfBT commented:

It was noted that only a few schools of those inspected were actually putting AfL into practice and found that of those that were implementing it, many were not introducing it in accordance with the way advocates suggested was most desirable. That is why the practical feasibility of the approach has become an issue requiring further research in relation to this way of understanding assessment.¹⁶

In short, AfL—as conceived by Black and Wiliam—has been a dead letter in most schools for a long time, but no one wants to admit it. The Wikipedia entry for AfL is revealing: it refers to the QCA and the DCSF, both of which were disbanded by the Coalition in 2010, yet in six years, no one has bothered to update the entry.¹⁷

The huge distance between faculties of education and the realities in the average comprehensive school has seldom been fully understood by policy-makers. As an academic discipline, education has long been a global enterprise, and one suspects that many of our professors are more at home on the international lecture circuit than in a school classroom. AfL is an international phenomenon and it has generated an enormous amount of literature. So far as we have been able to determine, little if any of the 'strong

14 declara, op cit

15 Willingham, D T (2009) *Why don't students like school: A cognitive scientist answers questions about how the mind works and what it means for the classroom*, John Wiley & Sons.

16 Flórez, op cit, p 14

17 https://en.wikipedia.org/wiki/Assessment_for_learning Accessed 30/01/2017

research evidence' cited by Coe has been arrived at through rigorous quantitative analysis. The CfBT report—which is generally very supportive of AfL—allows that

There is only one quantitative study that has been conducted which was clearly and completely centred on studying the effect of AfL on student outcomes. This produced a significant, but modest, mean effect size of 0.32 in favour of AfL as being responsible for improving students' results in externally mandated examinations. It must be mentioned, however, that this study has some methodological problems, explicitly recognised by their authors. These are related to the diversity of control groups they considered and the variety of tests included for measuring students' achievement. All this affects the robustness of comparisons within the study.¹⁸

Although AfL may have produced some beneficial effects in schools where staff are committed to its principles, attempts to impose it on all of our schools have contributed enormously to teacher workload and stress.

Yet even the best of ideas will fail when teachers are not convinced of their value. As an example, synthetic phonics is very simple and straight-forward. We have very robust and reliable standardised tests of children's early reading development and we know that many schools employing synthetic phonics have all but eliminated reading failure. This method was first developed in England by primary school teachers such as Sue Lloyd and Ruth Miskin, who weren't afraid to challenge the official orthodoxy laid down in the 1975 Bullock Report.¹⁹ Although teaching pupils to sound out words was not new, it was unheard of for schools to rely exclusively upon decoding graphemes to phonemes for beginning reading instruction. Throughout the Anglophone world, children were taught recognise whole words by sight, and to predict unknown words using different 'cues' such as context, syntax and pictures.

By 2005, the evidence for synthetic phonics was too powerful to be ignored, and in 2006 the Rose Review set in train its implementation. Unfortunately, the government decided to ignore the excellent synthetic phonics programmes already in place—ones which had been developed by teachers and published commercially—and develop its own programme, *Letters and Sounds*. The only advocate of synthetic phonics involved was Jennifer Chew, who has an encyclopedic knowledge of reading research, but no

18 Flórez, op cit, p 17

19 Burkard, T (1999) *The End of Illiteracy? The Holy Grail of Clackmannanshire*, Centre for Policy Studies, London

experience working in primary schools. Coming hard on the heels of the 1998 National Literacy Strategy, which embodied the 'eclectic' approach which had long been standard in English-speaking countries, *Letters and Sounds* would have had a difficult reception irrespective of its merits.

Yet even when taught with conviction, it failed to teach *all* children to read. This was entirely predictable: around 20% to 25% of all children need significant amounts of over-learning before they can become skilled at decoding letters to words. Pupils with learning difficulties need quite substantial amounts of additional help. The singular merit of synthetic phonics was that these children were identified in the first weeks of school and teachers were able to provide that help before they fell behind and lost confidence. Unfortunately, *Letters and Sounds* was a one-size-fits-all programme which made no provision for pupils who needed more help. Rather than providing additional phonics lessons, sceptical teachers merely concluded that synthetic phonics had been over-hyped and that these children needed a more traditional eclectic approach.

By contrast, Professor Tommy MacKay used a bottom-up approach in West Dunbartonshire, a local authority consisting largely of high-rise slums and second only to Glasgow in terms of social disadvantage. Over a period of 13 years, he worked closely with all 46 primary schools, winning teachers' confidence by convincing them that disadvantaged children could achieve the same standards of behaviour as their more fortunate peers. He then introduced commercial synthetic phonics programmes developed by teachers in England. By 2007, all children in West Dunbartonshire were leaving primary school with a reading age of at least 9 ½ years²⁰—which is adequate to read and understand most newspapers. Reading scores above this level are largely tests of verbal intelligence.

In England, we still have very little idea how well our young children can read; our 7+ National Curriculum Tests are administered by teachers who have a vital stake in the outcome. Unfortunately, the introduction of *Letters and Sounds* sent a clear message to teachers: don't use your initiative.

Ofsted

School inspections are a peculiarly English innovation, but unlike railroads, tanks and televisions, they have not been copied in other countries. There is something a little

20 Burkard, T (Spring 2007) A World First for Scotland, *Salisbury Review* 25/3, p 13

unlikely in the idea that a team of teachers who—for one reason or another, have decided they've had enough of teaching—can observe a school in a formal, Potemkin-village inspection, and tell us how good it is and how it could improve.

Like teachers, Ofsted inspectors vary considerably in their background and their views about education. Since Ofsted's inception, they have had to respond to quite profound shifts in education policy, and equally diverse leadership. The traditional views of Chief Inspectors Chris Woodhead and Michael Wilshaw stand in stark contrast to those of Christine Gilbert, for whom education was just as much a matter of social and emotional development as of academic excellence. Inspectors would be less than human if they did not regard these changes sceptically and rely to a large extent upon their own values and judge schools and teachers accordingly.

Like all the other principals in the 30-year saga of failed reform, we may be sure that they honestly want to do what they think is best for our children. However much they might try to faithfully implement the policy of the day, their biases will inevitably lead to differing interpretations and emphases in their judgements. Schools facing an Ofsted inspection have no way of predicting what these biases will be, and hence may feel the need to provide evidence that will satisfy a range of expectations. Of one thing we may be reasonably certain—on the day of inspection, the tableau presented by the school is highly unlikely to resemble what happens at any other time. Now that schools are likely to be inspected with less than 24 hours notice, they have to be prepared to slip into 'inspection mode' at any time—further increasing general stress levels in struggling schools.

In an explicit recognition of the disruption caused by Ofsted, schools that have good exam results or test scores are subject to less frequent and less obtrusive inspections. However imperfect these high-stakes exams may be, they still produce far better evidence of pupils' attainment than can possibly be had from a brief observation of a few classrooms and an examination of various artefacts. Quite rightly, government policy now recognises that we should not be too prescriptive about what teachers do, so long as their pupils are making good academic progress. Unfortunately, existing measures only test achievement at 7+, 11+ and 16+ and do not identify specific shortcomings when test results are below expectations: hence the perceived need for inspections to identify weaknesses.

However, the current protocols favoured by Ofsted inspectors are not supported by research. The obsession with marking pupils' workbooks is based upon findings that feedback correlates highly with pupil achievement. Undoubtedly, immediate feedback—

such as when pupils use worked examples for solving new problems—reinforces success, and enables pupils to correct mistakes before they become embedded. However, generic comments written in pupils' workbooks by overworked teachers are unlikely to be read by pupils, let alone used to good effect. In a review of the literature, Coe concludes that

Inferring the quality of teaching and learning from looking at artefacts such as student work, marking or lesson plans, or from assessing teacher portfolios, is not currently supported by research as valid.²¹

Teachers don't need the benefit of this research to tell them that marking pupils' workbooks is a waste of time—and what's more, one which prevents them from planning and delivering lessons which really would impact upon learning. Although Ofsted has recently been reviewing this evidence, schools still don't feel that it's safe to ignore the elaborate marking protocols that have been extensively promoted in recent years.

Lesson observations can do little more than reveal the obvious; when pupils are manifestly ignoring their teachers, we can be fairly certain that they aren't learning much. Yet as Coe wryly observes, positive behaviours such as busyness or engagement are poor proxies for learning. Unsurprisingly, the use of observations by external examiners to determine teacher effectiveness has little support from research.²²

Other than Ofsted's self-evaluation, there is little evidence to suggest that inspections lead to improved exam results. One of the biggest surveys, involving 2,700 inspections over a period of five years, revealed that the opposite is true:

It is found that there exists no evidence that the occurrence of an Ofsted visit has beneficial effects on the exam performance outcome of the school following the inspection. Indeed, the results show a small but well-determined negative direct effect on exam results: Ofsted inspections seem to affect adversely student performance in the year of the visit.²³

For better or worse, governments in England have long relied upon HMI or local authority school inspectorates to ensure accountability. It is certain that schools must be

21 Coe, R, Aloisi, C, Higgins, S, & Major, L E (2014) *What makes great teaching? Review of the underpinning research*, Centre for Evaluation and Monitoring, University of Durham, p 44

22 *ibid*, pp 30-31

23 Rosenthal, L (2004) Do school inspections improve school quality? *Economics of Education Review* 23 (2).

accountable: almost all maintained secondary schools receive well in excess of £5 million annually from the taxpayer. Schools must stay within the law, especially in respect to their financial arrangements and governance, so this will remain a necessary and legitimate remit for Ofsted or local authorities. Academic standards are another matter altogether: since its inception in 1992, Ofsted has been the major instrument by which government has initiated change and attempted to improve academic standards. If anything, academic achievement—especially in the STEM subjects—may have dropped.²⁴

One need not be a Burkean to be wary of upsetting established arrangements, and in any case it is unlikely that any government would abandon existing measures designed to improve inadequate schools until a superior alternative had demonstrated its worth. However, the need for a new approach is manifest.

Objective tests of knowledge and understanding

In 2006, Sir Kenneth Robinson delivered a TED lecture lambasting 'factory schools' and England's presumed culture of tests. To date, it has been viewed 42 million times,²⁵ making it the most popular TED talk ever. In 2015 Tristram Hunt, then the Labour education shadow, quoted Sir Ken approvingly:

The emphasis on testing comes at the expense of teaching children how to employ their natural creativity and entrepreneurial talents – the precise talents that might insulate them against the unpredictability of the future in all parts of the world.²⁶

Hunt, seemingly oblivious of the failure of AfL, goes on to reiterate Robinson's vision of personalised learning. Yet for anyone of a remotely liberal disposition, the arguments against testing might seem compelling. Stobart opposes tests on the grounds that

Assessment, in the form of tests and examinations, is a powerful activity which shapes how societies, groups and individuals understand themselves.²⁷

24 Harris, M (2008) Education Briefing Book: IoD Policy Paper, Institute of Directors, London

25 https://www.ted.com/talks/ken_robinson_says_schools_kill_creativity Accessed 30/01/2017

26 <https://www.theguardian.com/books/2015/apr/23/creative-schools-revolutionising-education-from-the-ground-up-ken-robinson-lou-aronica-review> Accessed 30/01/2017

27 Stobart, G (2008) *Testing Times: The uses and abuses of assessment*, Routledge, Abingdon, p 1

However, exactly the same criticism can be made about compulsory education; for most parents, there is no practical alternative to sending your child to a school run by the state. In the postwar era, an elite of comprised of professional educators and political activists has profoundly altered the nature of our schools. Some of these changes were overdue: the cane was a relic of more barbarous times, and bullying was often tolerated as a useful preparation for adult life.

On the other hand, the post-modern assault on our common intellectual and cultural heritage lacked popular support. Until quite recently, even the left understood that our radical thinkers, intellectuals and innovators were worth studying. They knew that literature was a powerful stimulant of social change in Victorian Britain. They valued the scientific and technological advances which have freed the bulk of mankind from hunger and a wide range of deadly diseases. Now—as always—those on the frontiers of progress are standing on the shoulders of giants. The cognitive sciences have proved beyond a doubt that the 'creativity' that Sir Ken cherishes cannot emerge in an individual who lacks a wide range of knowledge and a thorough understanding of the subjects concerned.²⁸

In fact, Stobart's objection to tests underscores their value: they are the most powerful means by which a democratic society can thwart the ambitions of an entryist elite. Since 2010, ministers have understood this, and the revised SATs, GCSEs and A-levels have placed a much greater emphasis on understanding and knowledge.

The next step in this revolution is the creation of annual tests in academic subjects. Optional SATs in Years 4 and 5 are a reflection of this need, although for the most part they lack the academic rigour of the new GCSEs and A-levels. However, there are a lot of objections to annual testing, and these must be answered.

Stress:

The most emotive and potentially damaging argument against tests is that the stress they create contributes to the perceived growth in children's mental health problems. A 2015 article in the *Independent* cited a study commissioned by the NUT, which supposedly found that

²⁸ Baer, J (2012) Domain specificity and the limits of creativity theory, *The Journal of Creative Behavior*, 46(1), 16-29.

Pupils at Britain's schools are suffering from a rise in self-harm, anorexia, and other mental health issues because of a focus on "constant testing" . . .²⁹

The new evidence consists entirely of subjective reports by teachers. The main outside source quoted is ChildLine, which reported "...a 200 per cent increase in counselling sessions related to exam stress between 2012-13 and 2013-14". However, problems in school account for only 5% of the of ChildLine sessions for 12-15 year-olds, and 6% for children younger than 12 years. In addition to exam pressures, these sessions dealt with "concerns about performance, not coping with workload, dislikes school, new school worries, problems with teacher, truancy etc".³⁰

No doubt a few of the children who reported exam stress have harmed themselves or had mental health issues, although we have no idea what other factors might have contributed to this. However, the *Independent* report is highly misleading: ChildLine counsels less than 300,000 children each year out of a total of approximately 11,000,000 in the UK. Of these, about 5.5% were worried about their schools. Of these 16,500 unfortunates, we have no idea how many were specifically upset about exams or some other aspect of their schooling. In England (which has 85% of the total UK population) pupils only sit exams at ages 11, 16 and possibly 18, so it is unlikely that more than one out of every 5,000 UK children were so upset by the prospect of a blank examination paper that they needed counselling. Bearing in mind the stresses that we all face in life, the notion that 'exam anxiety' is equivalent to mental illness is almost an insult to children who have genuine mental health issues. Children who phone ChildLine may be upset and often have serious problems, but it is much less likely that significant numbers of them are actually mentally ill.

Nonetheless, when teachers have a negative attitude about testing, it would be surprising if this were not reflected in their pupils' attitudes. For this reason alone it would be a big mistake to force tests on schools where this kind of attitude prevails. The only way this negativity will be changed is when the doubters see what a massive difference systematic testing can make.

Critics of testing overlook the fact that it is the relative rarity of tests that makes them stressful. When they are a normal part of teaching and learning, they aren't stressful at all.

29 <http://www.independent.co.uk/news/uk/politics/over-focus-on-exams-causing-mental-health-problems-and-self-harm-among-pupils-study-finds-10368815.html> Accessed 30/01/2017

30 <https://www.nspcc.org.uk/globalassets/documents/annual-reports/childline-review-under-pressure.pdf> Accessed 30/01/2017

Far from compounding exam stress, routine testing can actually alleviate it. Pupils whose knowledge and understanding of their subjects has been confirmed by regular tests will have far more confidence when sitting a high-stakes exam.

Not so long ago, tests were routine in many independent schools; a primary school teacher who attended one on the assisted places scheme recalls:

Tests were constant, and in pretty much all subjects. I don't recall any fear or anxiety because they were pretty simple and if you didn't pass, then you had to revisit in order to catch up; it was normal, for everyone. What was excellent was the fact that I felt like I was getting better all the time, and it was a source of pride to be passing those tests too, constantly moving on to the next level. I wasn't set any targets though because it was just assumed that everyone would keep ploughing forward.³¹

The following remarks were provided by a science teacher in a school that serves a disadvantaged community, and they demonstrate the extent to which routine testing can transform schools:

When our department first introduced knowledge tests, there was a lot of grumbling—everyone thought it was just another crazy new idea, and the pupils would hate it. At first they did—even though we started slowly, just testing them on 5 definitions in the first week. However, inside of a few weeks, everyone was taking their fact sheets home to study, and they were really pleased with themselves when they got all (or almost all) of them right on the test. They enjoyed it far more than producing a page or two in their workbooks, mostly because they knew that they were learning much more than they ever had before. Soon they were learning 20 definitions every week—the harder we made it, the more they rose to the challenge.

For example, when my Year 10s arrived for their test today, they didn't look at all stressed—just very excited. They were testing each other on equations outside in the corridor. You don't get that buzz if you don't test kids. Even our bottom sets are enthusiastic—they are saying that science is their favourite subject and are performing unbelievably well compared to their own previous performance. They are behaving in every way like a much higher set.

31 <https://thequirkyteacher.wordpress.com/2016/11/19/keep-it-simple/comment-page-1/#comment-3702> Accessed 30/01/2017

Now, the department is almost fully on board, and other departments are considering the switch. Not only has our pupils' behaviour and desire to learn improved enormously, but weekly tests are a much more honest (and far less time-consuming) way of providing 'proof of progress'. And we know that when this system feeds through to GCSEs, we will be meeting or exceeding the most ambitious DfE targets for 'closing the gap'.³²

The tests being introduced at this school are carefully constructed to ensure that all pupils are fully prepared for GCSEs. They are heavily weighted towards knowledge and understanding. Although there are opportunities for the most able pupils to demonstrate higher-order skills, the only differentiation is by outcome. And herein lies a clue to the eagerness of these pupils: all of the learning tasks they are set are—unlike AfL's 'success criteria'—concrete. Pupils understand exactly what they need to learn.

Do tests distort the curriculum?

Normally, this is quite a legitimate criticism. Goodhart's Law states that "When a measure becomes a target, it ceases to be a good measure".³³ High stakes tests force teachers to concentrate on what is being tested—often to the exclusion of everything else—and 'gaming the system' becomes an art form discussed incessantly on teachers' blogs.

What we are proposing neatly avoids this objection: annual tests will include everything that has been taught in each subject in the preceding year. With the old theory tests for a driving license devised by the DVLA, each individual test paper selected 35 items from a question bank of 800. There was no way to prepare for the test other than learning the correct answers to all 800 questions. Our annual tests will work in very much the same way.

Schools that opt for **Ofsted-exempt School** status will understand the need to design their own tests for use throughout the school year. Two recent books—*Making it Stick*³⁴ and *The Battle Hymn of the Tiger Teachers*³⁵—explain both the theoretical and practical aspects involved. Ironically, once teachers realise that their pupils actually remember what they've been taught, they will have the confidence to stray beyond the boundaries of the curriculum and introduce enriching activities and new subject matter.

32 Because this is still a sensitive issue in this teacher's school, we cannot divulge the source.

33 Quoted in Stobart, op cit, p 116

34 Brown, PC; Roediger, HL & McDaniel, MA (2014) *Making it Stick: The Science of Successful Learning*

35 Birbalsingh, K (2016) *The Battle Hymn of the Tiger Teachers: The Michaela Way*

Testing declarative and procedural knowledge is not technically problematic:

Designing high-stakes tests which affect pupils' future educational opportunities demands the highest possible standards in terms of test design--which is why the continued use of essay questions is highly questionable, as is discussed below. Items designed to test higher-order skills can vary significantly in difficulty and must be weighted carefully if tests are to be fair.

Although our annual tests will unquestionably require professional input, the great majority of questions will not be problematic. Validity will not really be an issue, as items will be selected specifically for their relevance to a specified exam. However, the main point is that our annual tests are not high-stakes tests for pupils, although they undoubtedly will be used to ensure that slower pupils get additional help. As such, we need not be overly concerned with weighting questions. After a year or two, the data generated by test results can be used to ensure that individual test papers are equivalent in difficulty, but this need not be an initial concern. In the sense that the annual tests are high-stakes tests, the only consequences are for schools and teachers; variability in test items would not significantly affect reliability when 25+ test papers are averaged. Nor would there be any reliability issues when tests are used to assess the effectiveness of an initiative or teaching strategy.

The paradox of teaching children to think

The most controversial feature of our proposal is that tests must be completely objective and they should focus primarily on knowledge and understanding. This will certainly raise the most strenuous objections from all quarters of the education establishment, who will object that we are trying to turn our children into parrots, along with the predictable laments about 'factory schools'. Scholarly articles on assessment stress the need for varied kinds of assessment to measure 'deep' learning; it was exactly this kind of thinking that led to the introduction of AfL.

We've been here before. Prior to the introduction of synthetic phonics, the received wisdom was that phonics taught children to 'bark at print'. However, barking at print is a much better place to start than not being able to read words at all—which was the alternative for about 20% of our 6-year-olds.

Likewise, knowledge and understanding—valuable in their own right—are the raw material for higher-order skills and problem-solving. After all, you can't think without something to think about. For example: given the task of finding the area of a circle with a given diameter, you must know that the radius is half of the diameter, and that the answer will be equal to the square of the radius times pi. You need to know how to divide by 2 (or to multiply by 1/2 or 0.5), how to square a number, and how to handle decimal points when multiplying or dividing (and this assumes that you have automatic recall of number bonds and fluency in using efficient algorithms). And if you are asked to express the answer to the nearest square centimetre, you must know that there are 10,000 square centimetres in a square metre, and you must know how to round up or down.

That's a lot of knowledge needed for one simple problem—and if any one of the elements needed to solve the problem is missing, you will get bogged down and probably get the wrong answer. With this in mind, it's easy to see why exam questions in maths ask you to 'show your working'—it's taken for granted that many pupils will not have mastered all of the facts and sub-skills needed to solve a given question, and that they should be allowed partial credit if they've understood enough of them to have a reasonable stab at it.

Unfortunately, this is entirely the wrong way to go about building "thinking skills". In an excellent series of blog posts on "Making kids cleverer"³⁶, David Didau reviews the literature from the cognitive sciences. In the first instance, there is now no dispute that limitations in working memory are the bottleneck in terms of synthesising new information with existing knowledge. In other words, our ability to solve problems and think critically is limited by the capacity of working memory. Unfortunately, there does not appear to be anything teachers can do to improve this capacity—what you are born with is all you will get.

However, Didau explains that this does not mean that we can't improve pupils' ability to solve problems:

1. Extended networks of ideas and information (schemas) take up the same amount of space in working memory as single isolated facts so the more we know about a subject the more space we have to pay attention to novel ideas and interesting combinations of ideas.
2. Through practice we can automatise various procedural knowledge so that it

36 <http://www.learningspy.co.uk/psychology/making-kids-cleverer/> Accessed 30/01/2017

becomes background knowledge. When a skill or a fact has been automatised, we are no longer consciously aware of it and it takes up very little space in working memory allowing us to concentrate on things we haven't yet mastered.³⁷

The pupil who has automatised all of the procedures to determine the area of a circle in the above exercise will then have little trouble learning how to find the volume of a cylinder or a cone. One who has not will struggle—and lower ability pupils will very likely fail altogether.

Didau uses Cattell's distinction between *fluid intelligence* and *crystallised intelligence*³⁸ to explore the implications for teachers. Although the capacity of one's working memory correlates strongly with fluid intelligence (or 'g') and does not improve with training, there does not appear to be any limit to our crystallised intelligence—which is our ability to learn and utilise procedural and factual knowledge stored in long-term memory. This, we believe, is why pupils (and especially disadvantaged pupils) respond so eagerly to rote-learning tasks: for once, they know they are learning something useful and interesting.

And even more importantly, the more we learn, the easier it is to learn even more:

The real benefits become clear when we understand that by improving crystallised intelligence we can 'hack' our fluid intelligence. That is to say, we can use what we've stored in long-term memory to compensate for deficits in working memory. When we store information in long-term memory it gets organised into schemas – interconnected webs – which mean that when we retrieve one item we also bring with it all the information we've stored in the same schema.³⁹

The more connections new information makes with the schemata already in our long term memories, the more each of these schemata are enriched, extended and connected. This in turn allows much more elaborate representations to be drawn up from long-term memory, thus allowing us to solve more difficult problems and to gather more sophisticated insights. In other words, we enter a virtuous cycle identical to Stanovich's

37 ibid

38 Cattell, R B (1961) "Fluid and Crystallized Intelligence"

39 Didau, op cit

*Matthew Effect.*⁴⁰ And insofar as Sir Kenneth Robinson's beloved 'creativity' is concerned, we have a far richer palette of colours to choose from.

None of this is to suggest that the annual tests for our **Ofsted-exempt Schools** should exclude challenging questions that test higher order skills. Not only do pupils relish a challenge, but we need to demonstrate that our children are indeed developing higher-order skills much more effectively than they are with approaches such as AfL, which place far less emphasis on knowledge.

Exams: dethroning the 'essay question':

Unfortunately, it is impossible to assess writing skills objectively. Attempts to do so have seriously compromised our high-stakes exams, forcing pupils to practice writing formulaic responses that bear very little resemblance to effective writing. If it is possible to teach children to write well, we have not yet discovered how. At best, we can teach them basic spelling, punctuation and grammar. Almost certainly, pupils' writing improves to the degree that they have mastered the content of their exposition and can recall and reproduce coherent and well-developed schemata relating to the subject of the essay.

Leaving aside the subjectivity and inefficiency of essay questions,⁴¹ it is clear that using them as the central element of testing in many subjects has not improved our children's writing, as this *Guardian* article demonstrates:

...after marking GCSE exam scripts for a major UK examining board for the past two weeks, I can honestly say that not only are standards dropping, but also they are unbelievably low... In relation to the GCSE candidates' general standard of writing, as a part-time lecturer at a university, I had already become aware that many undergraduate students had abysmal reading and writing skills. However, even that did not prepare me for the written skills of your average GCSE candidate. The handwriting, most of the time, resembled that of a five-year-old toddler or a drunk (grotesquely simple or an illegible scrawl). A lack of basic punctuation, such as full stops, commas, capital letters etc., was commonplace. There were countless inarticulate, immature sentences, which did not make any sense to the reader.⁴²

40 Stanovich, K E (1986) Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy, *Reading research quarterly*, pp 360-407

41 Burkard, T (2009) *Ticking the Right Boxes*, Centre for Policy Studies, London

42 *The Guardian*, 25 August 2005

Even when technical writing skills have been mastered, Fordham warns that "a pupil who knows little produces a highly formulaic response to a question, falling back on generic structures and stock phrases to make up for the fact that he does not have a sufficiently powerful network of knowledge to activate in answering a question."⁴³

We suggest that ministers seriously reconsider the place of essay questions in GCSEs. At very least, no more than 25% of the possible marks should be allotted to essay questions in humanities, and there is a good case for abandoning them altogether for STEM subjects. In Year 11, pupils spend much of their time practising old exam questions, where learning how to make the most out of exam boards' mark schemes takes precedence over the ability to write cogent and persuasive essays—let alone learning more about the subject being tested. The use of rigid frames to shape pupils' writing ability is rather like using paint-by-numbers exercises to foster artistic ability. Insofar as essay questions will inevitably remain a part of post-16 exams, ministers should seriously consider a pilot of Adaptive Comparative Judgement, which produces reliable results without forcing pupils to produce formulaic essays.⁴⁴

Were it not for the fact that they have long been considered the gold standard for evaluation of performance in elite schools, the above considerations would have long since dictated much more extensive use of alternate means of assessment. Ministers should also consider the following:

1. When compared to essay tests that are adequately marked, multiple choice tests are 12 times more efficient in terms of the time taken to sit a test, and over 7,000 times more efficient in terms of marking time.⁴⁵
2. Marking essays is very expensive—GCSEs and A-levels are now a major industry, costing the taxpayer well in excess of £ 1 billion each year.
3. When teachers moonlight as exam-markers they place a burden on their colleagues and their own effectiveness in the classroom is compromised.
4. The subjectivity of grading is highlighted by the increasing number of appeals—last year, 572,350 results were challenged and 90,650 were changed.⁴⁶

⁴³ <https://cliotcetera.com/2016/11/21/the-knowledge-party-in-my-head/comment-page-1/#comment-3217> Accessed 30/01/2017

⁴⁴ <http://www.cambridgeassessment.org.uk/Images/296241-the-reliability-of-adaptive-comparative-judgment.pdf> Accessed 30/01/2017

⁴⁵ Burkard (2009) op cit, p 6

⁴⁶ <http://www.bbc.co.uk/news/education-35054399> Accessed 30/01/2017

The way ahead:

The demise of AfL has holed Sir Ken's vision of personalised learning below the waterline, and many of our teachers are beginning to understand that higher-order skills are domain-dependent and cannot develop in the absence of extensive knowledge in that domain. There are no short-cuts—and attempts to take them have put our less-able children at a huge disadvantage, even to the extent of putting them off education altogether.

Michaela Community School in Wembley is one of the many schools that have already introduced a knowledge-rich curriculum combined with routine testing. It serves a highly disadvantaged community and about 90% of its pupils are BME, yet the performance and attitude of their pupils is nothing short of astounding. As measured on the National Group Reading Test, pupils who are behind in reading make up to five years' progress in their first year. Pupils are all acutely aware of their advantages, and they vastly prefer Michaela's demanding regime to the child-centred chaos they experienced in primary school.

Crucially, all teachers at Michaela work a normal nine-to-five--no one ever marks workbooks. True, lessons are almost constantly observed—mostly by outsiders who are eager to find out their secrets. Their 'open door' policy allows teachers to observe each other's lessons informally. Lessons are highly interactive; although the teachers are in total control, the pupils' desire to learn is palpable, even to the casual observer.

However, the last thing our educational system needs is a new initiative where every school is expected to duplicate Michaela's system. Rather, we need schools where teachers and leaders want to do even better--or even just to do things differently and get similar results. For instance, in the primary sector, we already have schools teaching the Core Knowledge curriculum—**Ofsted-exempt Schools** would give them the chance to demonstrate just what their pupils can achieve. We need more schools to demonstrate that the vast majority of SEND pupils in primary school respond just as well to a knowledge-rich curriculum as do pupils without special needs.

The **Ofsted-exempt School** programme will not produce instant results--after all, Michaela is only in its third year, and they won't have GCSE results until 2019. But from the reactions we've had to this idea, I think that even our top academy chains will find it difficult to hold on to their best teachers once they have the opportunity to run a school without any reference to Ofsted.

Yet even as the knowledge-centred curriculum gathers support, it would be a mistake to underestimate the influence of the education establishment and the persistence of the belief that teaching knowledge is obsolete in the digital age. Likewise, the tests that are being introduced by exam boards are a promising development, but it will be a hollow victory if tests supplement rather than replace the oppressive and unfair subjective measures now used to provide 'proof of progress' for Ofsted.

Above all, ministers need to understand that top-down reforms have failed. They need to reflect on the remarkable synthetic phonics revolution, which was started in England by Sue Lloyd, a primary school teacher who doesn't even have the benefit of a degree-level qualification. What she started overturned the conventional wisdom espoused by the massed ranks of professional educators as reflected in the 1975 Bullock Report and the 1998 National Literacy Strategy. She could not have done this if Ofsted had existed when she was developing the approach we now know as synthetic phonics. Ministers must understand that we will never get the teachers we need until we stop treating them as menials who must not only be told what to do, but how to do it.