Beyond the exam factory

alternatives to high-stakes testing

More Than A Score

First published 2017
by More Than A Score
c/o NAPE National Office,
Moulton College,
Moulton,
Northampton, NN3 7RR

Details of authors and sources, and acknowledgements, can be found at the end. We are particularly grateful to Tim Sanders, cartoonist at Common Knowledge, who generously agreed that we should re-use his drawings, as well as Ros Asquith for her kind permission to use her cartoon on page 95.

More Than A Score is a broad coalition established by the largest teachers' union the NUT (now part of the NEU), parents' organisations, academic researchers, and specialist associations for school subjects, primary schools and early education.

This publication arose from a *More Than A Score* seminar at Oxford University in March 2017.

Printed by Russell Press Ltd, Basford, Nottingham NG6 0BT

Contents

Foreword

Much of what the government asks primary schools to do reduces children to figures and numbers. This has a slow but steady, deadly effect on everyone - children, parents and teachers - even when we are totally against it. In the end, we arrive at thinking of whole children as a 'something out of 10'!

In fact, our children are complex and varied and changing. They are good at all sorts of different things in different ways and the job of education is to nurture and bring out as many of these abilities and tendencies as we can.

It is so important not to think of children as one type. Labelling so early on can lead to children being marked for the whole of their school career. I can remember a time when we were marked as one kind of child when we were 11 and divided into grammar and secondary streams. That was bad enough. Now it happens right from the start of school.

This book offers teachers, education staff and parents a different model. It explores ways of assessing children that put their development at the heart of the process. It results from the coming together of organisations who have combined their expertise and passion for education in the *More than a Score* coalition to persuade decision makers to listen to those who know and care.

The latest DfE review was an opportunity to reduce the testing burden but instead it has increased it and shifted the start of the data-crunching machine to children aged four. We know there's a better way. We hope this book will contribute to bringing about real change.

Mike Rosen

Beyond the exam factory:
alternatives to high-stakes testing

Introduction

Assessment should be a close companion to teaching. Its key purpose is to see whether the intentions of curriculum are bearing fruit in our students' learning and development. The relationship is subtle and respectful.

By contrast, heavy-handed assessment has a negative washback onto teaching and undermines the best intentions of any curriculum. If assessment is not designed to enhance learning, it is likely to debase it. This has been the English experience for the past 25 years.

This book aims to expose the workings of a system in which assessment has become harmful rather than supportive, and to explore better alternatives. It particularly focuses on primary schools, given the current crisis, but is relevant to other age groups and indeed draws positive examples from early years, secondary and post-16.

High-stakes assessment: the English model

The term 'high-stakes testing' originated in the USA. It refers to testing which has serious consequences for teachers and schools - testing which has exceeded its purpose of supporting learning. England has invented its own extreme 'high-stakes' system by linking assessment to 'league tables', Ofsted inspections, performance pay, 'naming and shaming' schools, and eventually school closure and forced academisation. This complex and punitive system of 'accountability' exacerbates any problems in the tests.

This was not accidental. Its origins can be found in the attempt by Margaret Thatcher's Government to impose a neoliberal ideology onto public services.

The aim was to place schools in competition with one another; some would expand and others would go to the wall - a crude business model. (In reality this proved impossible; instead, after 2000, schools were forced to become academies.) The Education Reform Act (1988) brought in national tests as part of a package designed to create a 'market': parental choice would have primacy over catchment areas, new types of school were established, and schools given financial autonomy. This was soon followed by Ofsted and the publication of league tables of results (1992).

The polarising effects of this business model of competition soon became evident. Some schools were imagined to be superior because their scores placed them higher in the local league, even if this was simply a reflection of their intake. Others, situated in poorer neighbourhoods, tended to lose students and especially from upwardly mobile families. These undersubscribed schools were then required to accept students that other schools had excluded, leading to a concentration of troubled and needy young people. As the system developed, the struggling schools were stigmatised as Failing (now 'Inadequate') and later forced to become Academies.

Schools came under pressure to improve test data at any cost. This resulted in excessive test preparation, curriculum narrowing, 'gaming' with easier qualifications and the removal of less successful pupils. In other words, rather than improving children's education, the collateral damage of this top-down accountability system ended up damaging it.

Predictably, it was schools serving disadvantaged neighbourhoods that suffered most from comparisons between attainment data. Later, various methods of measuring progress rather than raw attainment were devised. 'Value added' was calculated by comparing final results with prior attainment.

However, although Contextual Value Added calculations were fairer, this did not resolve the problem: generally speaking, value-added calculations still reward schools serving professional families whilst schools in areas of unemployment with high levels of child poverty were damned. Even though the current system takes account of the normal progress of children with different starting points, it does not seriously reflect the extreme problems faced by children growing up in poverty - particularly in deindustrialised areas - or the more recent consequences of austerity politics.

Dubious claims

As this system became the norm, various beliefs became commonplace:

1) It is often claimed that high-stakes assessment raises standards - that without it teachers and heads would become complacent. It is true that scores in national tests and exams have had an upward trend over the years, but much of this is due to more intensive test preparation. Despite the rising test scores at age 11, England's performance in the international PISA assessment remains mediocre. One hypothesis is that SAT's requirements have become a distraction from longer-term cognitive development.

2) Assessment at age 11 is supposed to aid transition from primary to secondary education. Unfortunately, cramming for KS2 tests is now so intense that secondary schools no longer trust the results: children are frequently retested on entry to Y7.

3) It is argued that national testing safeguards disadvantaged young people and prevents underachievement. This is clearly not the case: indeed, schools in poorer areas are likely to be stigmatised and the attainment gap hasn't gone away. The combination of data-driven competition, 'naming and shaming' by Ofsted and forced academisation has exacerbated staff recruitment and

retention problems and made it more difficult to improve struggling schools in sustainable ways. It is also very hard for teachers to attune to students' interests at the same time as maximising scores in standardised tests.

4) External assessment is supposedly more reliable, yet scandals of poorly designed tests and flawed marking have become common. It now appears that children can pass or fail depending on how they draw a semicolon. Similar problems occur at GCSE and A Level.

5) Standardised assessment is supposed to help schools to monitor progress. In reality, the system is incoherent, with a tenuous relationship between each individual's scores at different ages. Rather than furnishing helpful data, the fear generated by high-stakes assessment has led many schools to test children every term or half-term - a neurotic response.

In so many ways, rather than the pressure from high-stakes testing driving up performance, their collateral damage undermines high-quality learning. We will return to some of these concerns later.

Michael Gove's revised National Curriculum: escalating the problem

Tests since 2016 have been based on the revision of the National Curriculum by the former Secretary of State for Education, Michael Gove. In many primary schools, this revised curriculum has brought assessment to the point of breakdown.

The new curriculum was supposed to make England a "global winner". A hundred academics warned Gove publicly that the new primary curriculum would be counterproductive. It is worth recalling their argument.

> 'The proposed curriculum consists of endless lists of spellings, facts and rules. This mountain of data will not develop children's ability to think, including problem-solving, critical understanding and creativity.

> Much of it demands too much too young. This will put pressure on teachers to rely on rote learning without understanding. Inappropriate demands will lead to failure and demoralisation.
>
> The learner is largely ignored. Little account is taken of children's potential interests and capacities, or that young children need to relate abstract ideas to their experience, lives and activity.'

The Government shrugged off this advice, throwing insults at the academic experts, but there is undeniable evidence since of:

- more teaching-to-the-test
- enormous stress on children, along with a fear of failure
- a narrowing of the curriculum
- large numbers of children being labelled failures and moving on to secondary school anxious and demoralised, with extreme effects on disadvantaged students.

In 2016 the new tests failed almost half of children (47%) in at least one key subject (Reading, Writing, Mathematics). In Reading alone, one in three were judged to have 'not met expected standards'. In Maths, 30% were failed.

This was even more disastrous for children growing up in poverty. Two out of three children on free meals were failed in at least one test. 51% were failed in Reading, 46% in Maths, 41% in Writing.

The tests were modified a little in 2017 and pass marks adjusted, but 2 out of 5 children were still deemed to have failed in at least one subject: 3 out of 10 were failed in Reading, and 1 in 4 at Maths.

It is irresponsible to label children so negatively after seven years at primary school. This will lead to enduring problems of demoralisation, low self-esteem and defeatism during their secondary education and beyond.

It is not the children who have failed but the tests.

Is there an alternative?

The many detailed responses submitted to the House of Commons Select Committee's enquiry into primary assessment show that there is no shortage of ideas in circulation, simply a lack of political will to listen.

Around the same time (March 2017), the More Than A Score coalition held a seminar at Oxford University to share knowledge and experiences of alternative forms of assessment which were once widespread, and still are in many countries. This book shares the evidence and discussion from that event.

The aim of this book is to break through the restricted assumptions about assessment which have become commonplace, and create a wider sense of possibilities. It is not intended to be a definitive blueprint of how to replace the current system, but to serve as a resource for re-imagining what assessment could look like.

The alternatives are not 'pie in the sky'; they were used in the past but were then eclipsed by national testing and fell into disuse. One of our purposes is to bring them back into circulation as viable and valuable alternatives.

This will not be easy, given the pressures which have been placed on teachers, reduced time spent on initial teacher education and CPD, and the spread of a culture of compliance in many schools. For the alternatives to flourish, we need a more collegial, less threatening working climate. In particular, the punitive system of control and surveillance which regulates schools and teachers' lives (league tables, Ofsted, an obsessive use of data, performance pay and enforced academisation) must be dismantled if professionalism is to

be restored. Nevertheless, there is sufficient knowledge, experience and co-operation in many local authorities and other inter-school networks to start to establish these alternatives to high-stakes testing. Teaching will become a far more attractive profession when some of the time currently devoted to test preparation and data is used for meaningful assessment and learning.

Part A begins by considering some key problems with tying assessment to 'accountability', including the damaging side-effects for the education system of an overreliance on numerical data. It considers the workings of Ofsted and overbearing styles of management in undermining teacher professionalism. It calls for collegiality and democracy, rather than top-down controls. This is not a denial of the need for schools to be responsive and responsible to parents and their students, but rather an argument that the current system does not achieve this.

It also questions the logic of control through data (or as it has become know, 'governance by numbers'), exposing the flawed assumptions of attainment measures and the evaluation of teachers and schools based upon them.

The next section focuses on the developmental needs of young children and the particular difficulties that arise when they are subjected to tests such as the proposed 'Baseline' test on four year olds.

Finally, we introduce some broad general proposals to replace the current system. These are not a definite plan, but intended to demonstrate that there are sensible and workable alternatives.

Part B examines various types of assessment in more detail. It is divided into various headings for convenience, but it is important to recognise that they overlap.

We begin with formative assessment (Assessment for Learning). A major aim is to show how formative assessment can help to empower the learner.

Diagnostic assessment is considered next, with examples relating to language development and early literacy. We contrast this with the Phonics Check, a simplistic parody of diagnostic testing.

The following section considers ways of strengthening assessment by teachers, particularly when it is used for summative purposes. Although formative assessment should be the most important, summative assessment should draw upon a wider observation of learning than is possible in sit-down tests and exams. This section shows how moderation links to professional development, with examples of flexible 'best fit' criteria to provide teachers and learners with a sense of progression. Of course, all of this will be undermined if the current system of interschool competition, 'naming and shaming' of schools, labelling by Ofsted, and so on, are not removed.

Skilful and sensitive observation is crucially important for interpreting and guiding the learning process, but also for summative assessment which is not simply paper-based.

Portfolios are an important way to collect evidence of learning and progression, as well as supporting transitions. Examples here show how they can also strengthen children's reflective involvement in their own learning.

Finally, the book introduces a variety of options for authentic or holistic assessment, including outcomes such as performance, presentations and exhibitions. Learners are challenged to present a design or the results of their enquiry to a panel of teachers, parents and specialists in the wider community.

Such methods also provide important ways to encourage curricular breadth and balance, and to recognise and evaluate creativity.

We hope our book will be a stimulus for discussion and a resource for change.

A message to parents

A frequent justification for England's complex accountability system (high-stakes testing, Ofsted, and systems of control within schools) is the need for parents to receive sufficient information on the quality of teaching and learning in their children's schools.

We would not question the rights of parents to be properly informed, but seriously doubt whether the current system provides that information. Genuine knowledge about what and how children are learning is stripped away when achievements are reduced to numbers. There are better ways, including descriptive reports, discussion of a student's progress, providing samples of writing, live performances and presentations. A portfolio can demonstrate the progress made over time, and parents can be invited into school to a display of children's work or to observe an activity.

There are better ways too of evaluating the quality of education than Ofsted inspections, which are far too short and superficial to do much more than check the data. In the past, before Local Authority budgets and staffing were run down, their advisers would monitor a school's development over time. Systems have been developed whereby a school's management team, in collaboration with the whole staff, students *and parents*, would scrutinise key issues, collect evidence, and discuss what improvements were needed. This process is strengthened by involving experienced heads and teachers from other schools as 'critical friends'.

Such methods are less punitive but more informative. They also strengthen parents' relationship with the school and their ability to support their children's learning.

We know that many parents are deeply upset by the pressures being placed on their children by the test regime, and concerned for their wellbeing and mental health. We have considerable evidence from research studies that children are becoming disaffected and alienated from learning, or focusing only on the grades rather than the pleasure and value of real learning.

In May 2016 parents in many parts of England protested by removing their children from school just before the tests, involving them instead in more creative and meaningful learning. We expect such protests will grow as it becomes clear that the government's response to the crisis is seriously inadequate.

More Than A Score was established as a broad coalition of the largest teachers' union, parent groups and academic researchers, as well as specialist associations for school subjects, primary schools and early education. It is our aim to support active opposition, especially from teachers and parents, to a test regime which is educationally and socially damaging and replace it by assessment which helps children to learn. We also know that many parents are school governors who are able to give active support to heads and teachers in opposing and changing the present system.

We hope this book will help its readers see beyond the confines of the present system, and that it will strengthen parents' ability and resolve to change schools for the better.


Rescue
ur
Schools
TUI

Part A Section 1

Assessment and the accountability machine

Teachers clearly have a responsibility to discuss students' progress with parents and carers and facilitate their support. Unfortunately, the elaborate machinery of 'accountability' in England creates a sense of fear, not trust. As Fred Inglis argues (extract 2), 'accountability' is not the same as responsibility, it is 'a pistol loaded with blame'.

Every year, a proportion of schools are placed before a firing squad, but many more are made to tremble. It is no wonder that the kneejerk response of some headteachers is to place unreasonable and unhelpful demands on teachers, often resulting in excessive workload, unnecessary activities designed to protect the school from being caught out by Ofsted, and feeding into the retention and recruitment crisis.

Our first two extracts come from articles published 16 years ago, focusing on Ofsted and on managerialism. They are challenging texts, but readers will immediately recognise their relevance to current problems of assessment. After all, Ofsted, along with bureaucratic and oppressive styles of management, are part of the 'high stakes' of the English assessment regime. Both extracts call for a sea change in power relations and a more supportive and trusting partnership between students, parents and teachers.

After that, we summarise some of the discontinuities found in the data. This is important because the system depends for its logic on the assumption of smooth linear progress as the norm, with deviations explained by the in/effectiveness of teachers and school leaders. The data presented here clearly exposes this myth.

Ofsted, inspection and the betrayal of democracy

Michael Fielding

On the self-serving smugness and insolent promiscuity of accountability: democratic deficit

Accountability has the feel of bureaucratic rationality about it. It lays down clear requirements for the accomplishment of certain tasks and outcomes. It tends to operate in hierarchical regimes where those who are accountable bear virtually all the weight of whatever is specified. Motivation tends to be extrinsic to the task in hand; the sustainability of the required workrate or outcomes has more to do with the *threat of penalties* than the fulfilment of internal satisfaction or moral obligation.

In sum, accountability tends to be a largely negative instrument of social and political control; it eschews any form of serious moral engagement in favour of a contractual or technical agreement; it operates most effectively within a psychological ambience of apprehension; and... it is particularly susceptible to the culture and practice of *blame*...

My second point about the lopsided nature of accountability arises from this last argument about inequities of power. Regarding parents as 'customers' sanctions an ever-expanding set of demands. *Customers* have and feel no obligation to play an active part in the accountability process other than to apportion blame or praise.

As Fred Inglis insists so elegantly and with such terrifying insight:

> 'Accountability' is, after all, not the same thing as responsibility, still less duty. It is a pistol loaded with blame to be fired at the heads of those who cannot answer charges.

The moral resonance of reciprocal responsibility

The discourse of *accountability* has no real place for the enduring mutuality of human engagement, but within an aspirant democracy *mutual responsibility* is central. Because responsibility is primarily a moral, not a technical or contractual notion, it elicits and requires a felt and binding mutuality that does not depend upon hierarchical arrangements. Motivation tends to be intrinsic to the general requirements of the practice and to the specific tasks.

In sum, responsibility tends to be a largely positive, morally resonant means of encouraging mutually supportive endeavour to which all parties feel reciprocally and interdependently committed... If failure occurs, the response is not to blame, but to require restitution and redoubled commitment within the context of appropriate support willingly given.

When we hold each other responsible, we do so in ways that tend to re-inforce reciprocal engagement. We foster dispositions and motivations that presume a human desire to do what is right and celebrate what is creative and joyful in each others' endeavours.

Relating means and ends: a personalist dialectic

In the *high performance* school - the kind of school that is embodied in the approach of OFSTED and of 'school effectiveness' - the personal is used for the sake of the functional. 'Community' is valued, but primarily for instrumental purposes within the context of the market place. Students and teachers are important but only in a derivative way, through their contribution, usually via high-stakes testing, to the public performance of the organisation...

In the *person-centred* school the functional is for the sake of, and expressive of, the personal. Its outcomes are widely and imaginatively conceived and its success is as satisfying morally and interpersonally as it is instrumentally. The arrangements we devise to enable schools to fulfil and demonstrate their democratic responsibilities towards the commmunities they serve will also be educative, engaging, inclusive and imaginative. This is one reason why Ofsted is anti-educational; it not only excludes the very things that are most important and enduring, but is also dispiritingly dull.

If the language of inspection is reduced and confined to school performance, then it must inevitably be a prison of its own myopia. If it embraces the language of education as a creative, exploratory process, then it has to find other ways of addressing the requirement of a communal, democratic responsibility that is honest in its intentions and forthright in its judgements.

A malediction upon management

Fred Inglis

Accountability is legal not moral. It is a principle of bureaucratic rationalisation. Evidence is produced not so much that duty has been done but that the documentation on hand codifies its doing.

Auditing is an act of policing. There is nothing necessarily forbidding about that. We all check up on others and ourselves. Certain corners of social life need more checking than others (shady corners, unfamiliar ones, ones where you don't know what's going on).

As Michael Power argues in his book *The Audit Society*:

> Pockets of doubt and checking may be created and institutionalized but surely not as an entire principle of social organization? The more one thinks about it, the more apparent it is that the imperative `never trust, always check' could not be a universalizable principle of social order: constant vigilance is somehow autodestructive.

Not only can one simply not have a society in which nothing is to be trusted, the development of auditing techniques, especially when audit as an idea so overreaches its originally financial limits, become first, ideological; second, pathological; third, venal. As soon as the ideology is accepted in the conversation of the culture, it breeds, as all parasites must, its own pathology. It insists on accountability as necessary where before there had been the inevitably messy give-and-take of human dealing.

'Accountability' is not the same thing as responsibility, still less duty. It is a pistol loaded with blame to be fired at the heads of those who cannot answer charges. The pistol is fired in public. Its lesson is that wounds shall be visibly inscribed on reputation. Pathology turns to psychosis - an unbroken cycle of checks as to quality and answerability from which there is never any escape.

The illusions of measuring linear progress

Reclaiming Schools

The current system of high-stakes testing is driven by the requirement to 'hold schools to account'. It depends heavily on attainment data, used to compare schools - 'governance by numbers'. This data drives crucial Ofsted decisions which can destroy schools and teachers' futures, and weighs more heavily than intelligent professional observation and advice. This summary of a substantial and growing body of research highlights the deep flaws and inconsistencies in the collection and use of data. (See www.reclaimingschools.org for further details)

The National Curriculum was originally designed as a set of ladders made up of 'levels', each one describing what a child knew or could do. Its *descriptions* gave teachers and learners a sense of progression, showing what improvement looked like. This was a huge step forward from the traditional "14th in the class" or "63%" in an exam.

Unfortunately its use for accountability purposes encouraged a panoply of illusions.

i) Because each level had a number, assessment was seen as *measuring* rather than describing. The assumption grew that more or less anything of significance could be measured, and conversely - what could *not be measured* was unimportant.

ii) Because each child had to be assigned a definitive level, the illusion took hold that a child would reach all aspects of the level at the same time.

iii) The numbers appeared to work like arithmetic: they were evenly spaced - indeed, could be divided into sub-levels. Consequently, the system began to

judge teachers by how long they took to move children from one to another ('expected progress').

iv) Numbers can also be averaged, so it seemed reasonable to judge a teacher's effectiveness by the mean score. Later, the initial score was subtracted from the final score and the averages used to judge 'effectiveness'.

This elaborate nonsense was based on a poor understanding of learning and human development, and on a poor understanding of measurement. Not everything can be measured in nature, and certainly not in human behaviour. We cannot measure happiness or ambition or kindness - or intelligence.

Even scientific laws of nature depend on circumstances: water does not boil at 100 Centigrade on a high mountain, and is rarely pure enough to freeze at 0. It is even harder to measure human learning in any meaningful way. It is uneven - learners meet barriers then leap forwards. Students can beaver away at details, then suddenly make a breakthrough.

Mastery is incomplete and contextual; assessment depends on how questions are asked, what tools are available, the student's worries or mood on the day. How well we learn does not depend simply on how much teaching we endure, but on what we already know and understand and whether we manage to connect the new to the old.

We don't even mean the same thing when we assess Mathematics or Reading on different occasions, and there may be little predictive relation between the different measures. Children who are good at counting at age five might not find it easy to master calculus at age 16. Different tests of 'Reading' measure quite different things: some score the pronunciation of single words out of

context, while others look for critical understanding of a text. There is no reason why progress should be smooth from one to another.

Finally, the allowances we make for personal circumstances might be misleading. It is easy enough to calculate an adjustment for Free School Meals, but this does not distinguish between the child of a graduate single mother who is temporarily unable to work and children growing up in areas of chronic unemployment.

It should be no surprise, then, that even the most elaborate calculations are misleading. The discontinuities revealed by various researchers show accountability to be threadbare.

Secondary teachers openly declare that they place no trust in judgements from KS2 SATs and tend to ignore them. Despite this, the numerical data can also endanger children's futures when it is used to determine which GCSE and A-level courses students are allow to study.

* * *

1) Early assessment is a very poor predictor of later achievement. It should not be used to judge the subsequent 'value added' by teachers or schools.

The most experienced Baseline test provider can only make correct predictions for 4 children out of 10 in terms of their likely attainment just two years later.

Teachers' assessment through observations provide richer information on a child's development, but are also misleading when converted from description to measurement. When the DfE converted the Early Years Foundation Stage profile into numbers, they concluded that only half of the variation in KS1

average points scores could be explained by the Early Years Profile: a child with an average score in foundation stage Reading was almost equally likely to receive a 1, 2a, 2b or 2c at KS1, and some received 3 or W [based on the system of sub-levels used at that time].

Despite this, children are too often placed in 'ability groups' as if we could accurately define their 'ability' or 'potential'. This too easily becomes a self-fulfilling prophecy.

2) The phonics check only reflects one aspect of reading. It relates poorly to reading for understanding.

The phonics check, with its non-words, is designed to ensure that teachers use the government-approved method for teaching literacy. Data analysis has shown that:

i) there is a surge in scores once the pass mark (32 / 40) is reached, with a dramatic increase in the percentage of children achieving each score - suggesting serious *inaccuracy* as children are squeezed from fail to pass

ii) the *youngest* children are twice as likely to be failed as the oldest, wrongly identifying many children as poor readers when they are simply not old enough

iii) there is a poor correspondence between phonics scores and KS1 results.

This poor match between phonics decoding and reading comprehension is well established by research. That is because reading for meaning involves a combination of abilities and knowledge (vocabulary, syntax, recognition of

common irregular words, engagement with texts, etc) and is more than decoding the letters.

The phonics pass rate (Y1) has risen dramatically since first introduced, but with no change in KS1 Reading scores for the same cohort of children a year later.

3) Children are not being given time to develop. Younger children are being put under extreme pressure and are at extra risk of being declared 'failing'.

A Freedom of Information request revealed that August born children were twice as likely to be failed on the Phonics Check. Now that the DfE regularly publish this data, it confirms that the youngest children are twice as likely to fail as the oldest at KS1 as well as in the phonics check.

4) Progress from one Key Stage to another is not smooth or linear. There is so much individual variation that it is unsound to use progress scores to judge teachers or schools.

The entire accountability system depends on an assumption that children normally make smooth linear progress from one stage to another, and therefore that teachers can be judged according to 'value added'.

The Education Datalab's expert statisticians have demonstrated that progression is extremely variable:

1. Only 55% of children get the KS2 level (age 11) which matches their KS1 levels (age 7)
2. Only a third of children at the average KS1 level (2B) go on to receive the average grade (C) at 16

3. Furthermore, of these children who do meet their predictions, the majority do so via a route that includes periods of slower and more rapid progress.

Variations in performance or progress can be caused by many factors, including poor health, poverty, family stress, having to move school, or (particularly among adolescents in areas of chronic unemployment) despair at limited prospects. It is a mistake to attribute these principally to a school's supposed ineffectiveness.

5) Individual schools can vary strongly from year to year.

Low or falling test results can result in unexpected Ofsted inspections, headteachers being pushed to resign, and forced academisation. This is inappropriate given that the stability of results from one year to the next is weak. (According to recent DfE statistics, the correlation between a school's results in one year and the next is only 0.6.) A less punitive, and more supportive and sustained, involvement of outside experts is needed for schools to achieve their best.

6) The accountability system invariably punishes schools serving poorer areas.

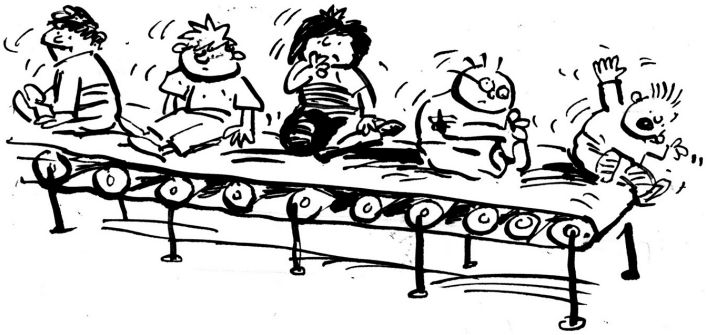
There is a general tendency for children growing up in poverty to make less progress than others. Those who begin with quite high attainment are often overtaken by more advantaged children who begin with lower attainment. Even judging schools through a 'value added' approach results in unfair judgements.

Family background is a major cause of variation (the impact of poverty, and conversely the advantages which better educated parents pass on to children). However, most of the variation is difficult to explain statistically (individual

temperament or attitude, focus on learning, general health, family breakdown, etc). Only a small amount (around 10-15%) of variation appears to be due to attending different schools.

Official data reveals that the best way for a secondary school to secure an Outstanding judgement is to mainly admit pupils with high prior attainment. Most judgements of *Requires Improvement* or *Inadequate* go to schools serving disadvantaged areas.

Even the carefully constructed Progress 8 is not a level playing field, despite government reassurances. Just a few troubled students in a class who fail to sit their GCSEs can wipe out the above-average progress of the rest of their class.



Section 2: Assessment and the developing child

The 2014 revision of the National Curriculum under Michael Gove was based on the premise that targets set for a particular age group in Singapore or Shanghai could simply be brought forward a year or two for England, and that teaching should be accelerated to reach these goals earlier. A century of knowledge about child development was discarded as irrelevant.

The first article in this section insists on the importance of understanding human development. Children are not learning machines.

The second considers some of the problems which emerged when Baseline tests were last introduced. It looks at the problems of predictive testing, and demonstrates the folly of current Government attempts to reintroduce Baseline tests as the starting line for 'holding primary schools accountable'.

The final article compares this simplistic and ill-conceived mode of assessment with the processes of observation, documentation and discussion developed in early years settings in Reggio Emilia, Italy. The article explores the two very different understandings of children which they reflect, and the different social values which underpin them.

Homo Sapiens 1.0: human development and policy construction

Pam Jarvis

Homo Sapiens appeared on earth approximately 200,000 years ago after a long period of evolution. One of its key traits is the immature condition in which it is born, taking a large percentage of its lifespan to become adult. The reason for this is presumed to be the complexity of the neuronal development that it needs to undertake - the ability not only to think abstractly but also to communicate these abstract thoughts to other people in a richly symbolic language.

The infant brain is in many ways quite different to that of a human adult, the key difference being its incomplete, malleable state. Babies' brains have far more neurons than adults', but far fewer connections. The early development of the brain involves an extensive neuronal connection program in response to environmental experiences. Those neurons that do not connect to others during this process shrivel and eventually die; such neural 'pruning' is an entirely natural process, the result of evolved human cognitive flexibility. This is essentially a 'nature via nurture' program: there is an in-built schedule for growth and development, but the direction in which the development occurs is directed by external stimuli. For example, there is a natural human ability to develop language, but the specific language spoken will depend upon the individual's environment.

Our current understanding of how the human brain constructs itself during the developmental period suggests that this happens via what is termed 'embedded mental representation', i.e. we incrementally memorise and co-ordinate our experiences. This generates an increasing ability to organise

thought, gradually resulting in the ability to manage incoming information and locate it within memory in increasingly sophisticated neural networks.

As children grow, there is an exponential development in their ability to organise cognition, in particular to focus attention without becoming distracted by the intrusion of non-relevant thoughts. This requires 'inhibitory behaviour', and the younger children are, the more difficult they find this; their thoughts are more susceptible to interference than those of adults due to the immature networks across which they travel. The more immature the network, the less capacity is available for incoming ideas to 'hang upon'; such capacity is gradually built as networks are constructed. A useful analogy is that it is easier to find something in a wardrobe where there are enough hangers for all the clothes. If we keep adding more clothes regardless of the lack of hangers, they will end up in a tangled muddle at the bottom and become very difficult to retrieve when we next try to find them.

Given the neuronal immaturity of young children, the DfE's quest of a 'baseline' assessment of the literacy and numeracy skills of four year olds in order to predict future progress seems quite ridiculous. Indeed, such plans have been pursued and dropped twice before, in 2002 and again in 2015. The reason given for testing such young children is to hold their teachers 'accountable'. Without empirical evidence, it is presumed that incremental progress can be accurately measured against this 'baseline'. The trouble is, literacy and numeracy are what are known as 'emergent' skills. In other words, human beings did not evolve to read or to count, but the abilities are emergent from other evolved competencies. These skills are certainly not part of the core neuronal developmental program when a human being is less than 60 months into the lifespan. Such a test is like judging a 'Bake Off' cake while

the ingredients are still being added to the bowl; as we all know, how a baked product eventually turns out is not simply the result of the amount of fat, sugar and eggs, but whether self raising or plain flour is added, how it is sieved into the mixture, and how the mixture is then beaten and baked. These variables act independently of one another; there can be no simple, predictable linear progression.

The way in which human beings naturally 'mix their ingredients' - or, from the perspective of an alternative analogy, 'boot' their cognitive system - in the first seven years of life, is through spontaneous, play-based interactions in which they independently interact with peers and adults. What this stage of development most crucially creates is not only the ability to *acquire* information, but also to *use it flexibly* - the basis of independent cognition. And in building relationships with others and sharing ideas children learn to engage socially and emotionally. This is a necessary stage in the development of communication skills and, given that human beings are highly social creatures, the basis for ongoing social and emotional health. Moreover, all human adult communication involves complex combinations of collaboration, cooperation and competition, skills that are developed through these early interactions.

Young children whose social, emotional and cognitive needs are poorly addressed are likely to develop poor stress-coping mechanisms. At the biological level, stress coping is mediated in mammalian creatures by the hormone cortisol, which functions as a 'thermostat' which turns up the 'alert' system when a stressful situation is encountered. The human stress response, like that of all mammalian creatures, is attuned to the need to either escape or

fight when threatened, and the release of cortisol begins a cascade of biological adjustments to make energy available for such response.

Children experiencing *ongoing* stress develop higher resting levels of cortisol, and the system takes longer to return to this base following a stressful event. Chronically heightened cortisol is not only linked with emotional disturbance but also impacts upon memory and learning. At the psychological level, ongoing stress runs a 'background program' in the mind, leaving less capacity to deal with incoming information. 'Too much too soon' approaches in early education that confuse and worry children contribute to cycles of stress and underachievement.

The results of contemporary care and education policies in the UK are illustrated in the deteriorating state of juvenile mental health. Statistics from *Young Minds* indicate that approximately one in ten children have a diagnosed mental health disorder. Nearly 80,000 children and young people suffer from severe depression, including 8,000 children aged under 10 years of age, while 20% of young people deliberately harm themselves. The New Economics Foundation found that the UK's 16–24-year-olds record the lowest levels of trust and belonging in Europe; two studies by UNICEF in 2007 and 2013 found British children to have a low sense of well-being compared to children in economically similar nations. Whilst there are many other factors that may negatively impact upon children's mental health, including cyber-bullying and family breakdown, the 'Too much too soon' approach to education and care is the element that we can most easily address. We need to adjust social policy for families and schools to match the natural developmental needs of human beings.

Baseline testing: science or fantasy?

Terry Wrigley

There's nothing hidden in your head
The Sorting Hat can't see,
So try me on and I will tell you
Where you ought to be.

The selection of children into houses at Hogwarts famously involves a magic 'sorting hat' - fiction, of course... unlike baseline tests in real schools. The Government's baseline tests at the start of Reception produce numerical data, so they have an aura of scientific accuracy. *Anything but!*

The following analysis focuses particularly on the tests designed by CEM as one of the three approved providers of Reception Baseline Assessment in September 2015. This is not because CEM are incompetent but rather the opposite: they were the most experienced providers. Their test was based on PIPS (Performance Indicators in Primary Schools), sold commercially to hundreds of schools in various countries and refined over more than a decade.

Predictive validity

CEM marketed its baseline tests as having 'excellent predictive validity'. Our subsequent investigations showed this to be a dubious claim. Perhaps their advertising unit really meant 'This is about as good as it gets!'

Other CEM documents showed a correlation of around 0.7 between the PIPS test and attainment two years later. 0.7 is often reckoned to be quite a strong correlation in the social sciences, but we have to ask questions of context and purpose. As a former civil engineer pointed out, when you're calibrating measuring instruments a correlation of 0.99 is disastrous: *bridges collapse*! A test

with a correlation of 0.7 purporting to predict cardiac arrest or alzheimers in the next two years would be unusable due to far too many false negatives and false positives.

Around the same time, another experienced statistician pointed out the need to square a correlation in order to estimate how much of an outcome can be predicted from an input measure. (This is because the formula for correlations involves a square root, which has to be reversed by squaring.) Squaring 0.7 results in 0.49, i.e. around half. In other words, only half of the outcome (eg a child's KS1 result) is predictable on the basis of the baseline test.

Further data came to light following a Freedom of Information request for a 'chances table'. This showed what proportion of children with each specific baseline score go on to reach different KS1 levels. It provides much more detailed information than a generalised correlation figure. The PIPS test made sound predictions of a KS1 sub-level for roughly 4 children out of 10. (Of course, levels or sub-levels are no longer used but that was the basis at the time and the problem of predictive validity remains the same however the data is presented.)

Predictions are quite good at the extremes, especially for children with unusually high initial scores at PIPS who tend to do very well at KS1. It was rather more problematic for low scorers, and many of these reached average attainment just two years later. For the vast majority however baseline scores were a poor guide. As an example, of children with a baseline score of 50 out of 100, 6% were graded W ('working towards') or 1 at KS1, 13% received 2C, 28% 2B, 32% 2A, and 21% ended up with level 3. [These details are for KS1 Reading, and are from a baseline test taken at the end rather than the start of Reception.]

The Government's intention was, and is, to link a baseline test undertaken in the first term of Reception with KS2 results nearly seven years later. The DfE resolutely refused to factor in the child's month of birth, a major factor for such young children. Tests had to be in English even for children speaking another language at home, which in itself makes nonsense of predictions. CEM's own researchers have also raised serious doubts about the predictive capacity of these tests in terms of emerging special educational needs.

Half a lesson learnt

As explained earlier, CEM have an established reputation based on predictive tests for various age groups. Judging by the data we saw from one of the other providers, it is shocking how lax the DfE were in vetting these providers. It seems that neither of the other two approved providers had any longitudinal data to underpin their bid.

The DfE soon realised they were presiding over a disaster and commissioned an independent evaluation from the Scottish Qualifications Authority. Unfortunately the DfE's conclusion was simply that data from the different providers could not be aligned. DfE officials failed to acknowledge the unreliability of any single provider's methods in themselves.

Testing very young children is particularly fraught with difficulties. A major problem is that test items are often borrowed from tests for older children. In other words, items originally designed to check whether a 7-year-old *has learnt* how to do something are used to determine whether a 4-year-old *will be able to learn it*. Absurdly inappropriate (and to many children, meaningless) test items were used such as:

> Say the word *parrot* without the P.

> I am going to sound out a word like a robot would say it: *p-i-n*. Can you tell me what word I have sounded out?

or the criterion:

> Can the child order and ascribe numbers up to 20 and add and subtract using single digit numbers?

Dividing up words into separate sounds is an artificial exercise which accompanies early literacy teaching: children are not born with the ability to divide meaningful words into phonemes. Whether a four-year-old can already manipulate numbers 1-20 reflects both the preparation received from parents or nursery, and the child's general level of maturation. To assume that such test items are reliably predictive of subsequent achievement is delusional.

Selection at four

Unfortunately such predictive assessment works, to an extent, as self-fulfilling prophecy. In other words, if you label all the low-scorers as "Low Ability" and treat them as such, and seat the high-scorers at a "High Ability" table, there is a stronger chance that they will reach the predicted attainment levels.

Sadly such practices - illegal and unthinkable in Scandinavian countries with their more child-centred educational traditions - have become all too common in England's test-driven primary schools. Some heads clearly believed the baseline tests would enable them to determine each child's *ability* and *potential* in the first few weeks of Reception.

It can be argued, of course, that baseline tests are only intended as a measure at whole-school level i.e. that by comparing the whole school's aggregate baseline scores with its aggregate KS2 scores, you can judge the school's "effectiveness". This may be possible but only if a school is large enough, if few children move in and out during the primary years, and if there are few

children with EAL, SEN or FSM entitlement. Regardless of the assumptions behind such judgements, there is an unavoidable problem: you can't do this without putting a label on each individual child.

Numerical data: the aura of science

Others have suggested that the Early Years Foundation Stage Profile should be used instead. The Profile is certainly a more holistic register of development than baseline tests focused on (proto-) literacy and numeracy, but once its descriptions are converted into numerical data, we encounter similar problems to formal baseline tests. The DfE tried this several years ago (DfE Research Report RR034). Among children with the midpoint score on reading in the EYFS profile, 23% went on to get W or L1 at KS1 Reading, 22% got 2C, 31% got 2B, 18% got 2A, and 7% L3. This is about as accurate as judging children by how clean their shoes are.

All of this forms part of a spurious search for certainty and the illusion that schools can be judged fairly by comparing value-added data. It forgets that, in general terms, children in poverty *tend* towards less progress even with very talented teachers. It is blind to the potential advantages that accrue from university-educated parents. It overlooks the difficulties of predicting progress for EAL children with variable exposure to English. It neglects the levels of mobility in and out of many urban schools. It ignores the variation between one year group and the next, exacerbated in smaller schools. It forgets that one sick child and an acrimonious divorce can upset the aggregate score of a one-form-entry school.

Unfortunately the subtleties have gone unnoticed by the largest headteachers' body. Clearly regarding an open challenge to KS2 tests as unachievable, the NAHT appear to have opted for Baseline as a fair basis for value-added

comparisons. It is sad that such a respected organisation should accept a high-stakes accountability system which is designed to put a proportion of their members in front of a firing squad each year, and identify their schools for academisation (or, increasingly, re-brokering into different chains). Is this evidence-based, or are there interests at work? A scrutiny of the NAHT's *Assessment Review Group* shows the direct involvement of two out of the three baseline providers - a case of vested interests. Similarly, schools Minister Nick Gibb, when asked to provide evidence for baseline testing, is forced to rely on one of the key providers. There is a circularity when key researchers are simultaneously policy advisers to government and recipients of major contracts.

Rather than impose this numerical data collection on Reception class teachers, their existing skills in observation exercised in the Early Years Profile should be honoured and enhanced.

Democratic Alternatives to Early Childhood Education Assessment

Guy Roberts-Holmes

Introduction

The Reggio Emilia schools continue to demonstrate that there are alternatives to the current utilitarian approach towards early childhood education. They were established by Loris Malaguzzi as an historical and situated response to Mussolini's fascism and the horrors of the Second World War. The innovative, visionary and democratic Reggio Emilia schools were based on Malaguzzi's political and ethical understanding that education should primarily be concerned with creating the democratic conditions in which children can become critical, creative and independent thinkers.

Malaguzzi articulated a culturally 'rich' image of the child necessary for such an emancipatory education: 'it is our moral duty to credit children, all children, with resources, possibilities and capacities that are much greater and much more universal than believed…and as the bearers of rights, values and competencies'.

This vision of the competent, sociable and democratic child (and teacher) with *rights* enables the complex narrative assessment known as *pedagogical documentation*. At its heart, pedagogical documentation is concerned with making learning and project work visible, and then subjecting them to democratic deliberation about meaning. This discussion potentially includes not only teachers, but parents, other citizens and children themselves.

Such democratic deliberation demands time, institutional support and an openness to joy, awe and the unexpected; not qualities much in evidence

today. It tunes in to the child's interests and activities that show what they *can* do. It is open to each child's unpredictability and diversity of potential. Today Malaguzzi's inspirational ideas are widely known and pedagogical documentation has become widespread, well beyond Reggio Emilia.

Secondly, and in stark contrast, Malaguzzi described the ways in which a negative, demeaning and disrespectful image constructs children as culturally 'poor' i.e. limited, inadequate and incompetent. Such children can be de-contextualised, categorized and regulated. They are 'indistinct children without qualities who stay where you put them and you can describe them as you wish, without gender or role or history'. All this can be achieved through the simplistic measurement of a basic utilitarian audit approach such as Reception Baseline Assessment, which is primarily concerned with the further regulation, governance and datafication of children, teachers and early education.

The Rich Child and Democratic Potentiality

Malaguzzi's image of the 'rich child' who has 'one hundred languages' with which to express themselves is *open to possibilities, potentialities and alternatives* and demands that the teacher engages in an intelligent and sensitive pedagogy of respectful listening. He stated that a teacher must

> vary, multiply, intensify, re-invent and re-listen to children's activities, behaviours, words and languages. Support and make use of their interests, their forms of learning, choosing and communicating.

Pedagogical documentation centres children's voices and their understandings by making children's learning 'visible' through sharing their creative arts, performance, photography, writing and play. Embarking upon connected

interdisciplinary project work chosen and led by young children with the teacher participating as a facilitator, enabler and co-constructor means 'being sensitive to the unpredictable results of children's investigation and research'. Pedagogical documentation can embrace diversity, uncertainty, contingency and unpredictable outcomes. Such democratic learning eschews reductionist notions of fixed learning objectives based upon normative targets, goals and expected predetermined outcomes.

One excellent example is 'The Crow Project', undertaken in a Swedish preschool with close links to Reggio Emilio . After a year's work on crows initiated by the children after a walk in the local woods, the children and the teachers sit together to discuss their collaborative learning in a process of democratic deliberation together. The documentation consists of children's (and teachers) drawings, paintings, photographs from the woods, research notes, plaster and papier mache models of crows. The teacher describes the children's learning apparent in the drawings:

> The children's drawings look almost like a film, which frame by frame, depicts how their curiosity shapes an ever more powerful relationship with the birds. The birds are no longer unknown entities flying overhead. They became the children's friends. And the children become extremely pleased with themselves when they discover the differences in their pictures over time.

As well as the drawings, notes were kept of children's questions, often unexpected ones (How do birds kiss? Can birds fart? Do birds think?) and the children were frequently involved in looking at the notes and their own drawings to discuss what they had learnt. This discussion, led by the children

themselves, without predetermined outcomes of any sort, itself generated more issues to explore:

> I wonder how they fly?
>
> I wonder what their feet look like?
>
> You know what I saw? That they could glide. They didn't need to flap their wings all the time. Sometimes their wings were perfectly still.

Individual children were asked to open the following day's meeting and their observations of nature, demonstrating the way in which the young children were trusted to make intelligent decisions about their own learning. This trust is at the heart of pedagogic documentation and leads to the children's democratic participation in the direction of their own learning. Drawing on the children's reflections helps the teacher understand things in a light other than the obvious. More specifically it enhances the children's participation in the group learning process by building on and tapping into their thoughts and actions.

Thus the process of pedagogic documentation is co-constructive, respectfully reciprocal and dialogic; it is reflective and generative of learning.

The Crow Project had a strong emphasis upon open-ended project work, listening to children and 'a strong belief in the unlimited potentiality of children'. It is focused upon the learning processes of participation, dialogue and imagination.

Baseline assessment and the culturally 'poor' incompetent child

Baseline Assessment, however, with its norm-based criteria which seek to govern and control through simplistic categories, numbers and linear outcomes is based upon the image of the culturally 'poor' child. Such a culturally 'poor' child depends on an 'exam factory' education where children

acquire and reproduce pre-determined knowledge as they are readied for the neo-liberal values of the marketplace. Such a child requires ever tighter regulation, control and governance if they are to be primary and secondary school-ready and able to compete in the DfE's so-called 'global race'.

Rather than being open to the child's possibilities as a thriving human being, baseline assessment has the tendency to judge children as a unit of potential 'human capital'. It represents 'a ridiculous simplification of knowledge and a robbing of meaning from individual histories' and involves a 'rush to categorise' and a sense of "does this child measure up" to a fixed and predefined norm.

Malaguzzi spoke about the dangers of a 'prophetic pedagogy' based upon prediction of what children should know and their required outcomes. A *prophetic pedagogy*:

> knows everything beforehand, knows everything that will happen, knows everything, does not have one uncertainty, is absolutely imperturbable…This is something so coarse, so cowardly, so humiliating of teachers' ingenuity, a complete humiliation for children's ingenuity and potential.

The reduction of a child's complexity to a single figure upon which predictions or 'prophecy' of subsequent development could be calculated is central to Reception baseline assessment. By judging four-year-old children with a battery of simple yes/no alternatives effectively negates, excludes and 'steals' children's 'one hundred languages'. In the following examples taken from one of the leading providers of Baseline Assessment, teachers had to judge children 'yes' or 'no'. Of course with such complex statements *it all*

depends upon the context, the social interaction and a vast range of other contigent factors, hence the absurdity of such decontextualized statements.

Literacy and Maths:

'Knows that print carries meaning and knows how to handle books'.

'Links sounds to letters, naming and sounding the letters of the alphabet.'

'Counts at least four objects.'

'Is confident to order and ascribe using numbers 1-20, add and subtract using single digit numbers.'

Each child was then given an overall score based on a crude calculation. This excluded contextual variables such as how long the child had been in school, their age, and whether or not English was their first language, as well as the young child's sociable and emotional variables. As one teacher commented:

> *I feel that the Baseline Assessment has to be completed too early in the year which means that teachers are madly trying to collect evidence, rather than concentrating on the welfare of their new pupils and helping to create a calm and relaxing environment which is vital for a positive start to their school life.*

This is ironic because the development of young children's well-being and learning dispositions are more important and reliable predictors of later academic achievement than simplistic indicators of literacy and maths acquired by the age of 4. The Baseline assessment is premised on the idea that a child's ability is fixed and their potential predetermined. Regarding such fixed notions of 'ability' Malaguzzi noted that 'such claims forget that plasticity is one of the central nervous system's characteristics… thus

declaring its susceptibility to experiences children live'. Children whose experience in the early years has instead supported emotional well-being, cognitive development and self-regulation during play may score less well on early academic tests, but evidence indicates that these children show higher achievement benefits in the longer term (see Whitebread and Bingham, 2013). Children in Finland, for example, begin formal schooling around three years later than England's Reception year (increasingly a misnomer, given current pressures). They follow active, play-based provision in their Kindergarten years; they go on to out-perform British children in later attainment (Bodrova et al. 2007).

As one headteacher told us:

> *I think doing any sort of reputable assessment of very young children is dodgy because the children are so young. You know if those children were in Denmark they wouldn't have had to pick up a pencil yet.*

Trying to assess children who had not yet sufficiently developed emotionally leads to a deficit model of assessment showing what they can't do as opposed to what they can do. This is evidenced by some of the classroom teachers' comments (below).

> *It's ridiculous. It's not a fair representation of children. Many young children are not yet confident enough to show their new teacher what they can do when put on the spot.*

Unfortunately, however, the strict DfE regulations meant that it had to be carried out regardless of whether or not the children had 'settled' in. No adjustment was allowed for the child's month of birth, despite the common sense, and evidence, that twelve months difference is enormous at this age

(for some, a quarter of their lives).

> *I did have children that were crying and I just couldn't get anything out of them at all because they were too upset to do anything, even when I left it till later on. Some children just refused or just weren't ready and I know they said you only assess them when they are ready, but some children, well, you got to the point where you had to assess them because it had to be done whether they were ready or not. And obviously then it is not accurate because they weren't at a stage when they wanted to say things.*

This leads not only to inaccurate data but is ethically inappropriate and potentially damaging for children's developing self confidence, self esteem and learner identity.

> *Some children looked at me and said "I can't read" when asked to read parts of the assessment. It was heartbreaking to see their reaction to it and I spent a lot of time reassuring children.*

Here baseline assessment had the inadvertent potential of demotivating and undermining young children's confidence in their abilities when they had only just started school. Baseline assessment thus established with pseudo-scientific rigour the setting of low expectations for particular groups of children, including summer-born children, EAL and SEN children. The potential for grouping and labelling children based on Baseline Assessment accountability data is a worrying development especially given its inaccuracy.

> *I don't think you should [use it to measure progress], I don't think you can, because they are children and they are not robots, not machines, they are children. You don't know what influences they have got from outside, what is going to happen in those seven years, so I think it is ridiculous.*

The variation between children's patterns of development and rates of progress means that any reductionist and simple correlation between Reception and Key Stage 2 is impossible:

> *Children's progress is going to be judged against how far they have gone in seven years. Now to my mind that is an almost impossible thing to do because you can't test children at 11 about the same things you were testing them at four. It just doesn't make sense.*

Refusing Baseline Assessment

Given the problems with baseline assessment it was not surprising that at least 3,000 head teachers refused to implement the 2015 trial. These included headteacher Dame Alison Peacock, who now leads the Chartered College of Teaching. In her primary school, she was able to establish alternative practices based on a rejection of the notion of 'fixed ability'.

> *Understanding children's thinking and their developing ideas through building and sustaining dialogue is an expert form of teaching that enables high challenge within a richly supportive environment. This is the beauty and the art of early years teaching that cannot be reduced to scores on a page, or to boxes on a tracking screen… we need to put assessment back in its box; thereby refusing temptation to place labels on children or their teachers.*

Refusing the simplistic and reductionist labelling of Baseline Assessment, Peacock's actions demonstrated that an apparently dominant, totalising and monolithic accountability regime was in fact contestable. Although at the time Baseline was voluntary, Peacock's and other headteachers' ethical and political *refusal* of baseline's dominant 'regime of truth' within a harshly punitive and disciplinary audit culture can be considered as a radical 'caring' for herself, the

staff and the children as she refused closure, regulation and labelling. Peacock's and other headteachers' refusal to engage with baseline assessment is a demonstration of a democratic accountability that is morally and politically situated.

Conclusion

At the time of writing the DfE is considering a further attempt at the introduction of a baseline assessment, this time using an individual tablet-based assessment. Such a reductionist approach to early childhood education will not only produce inaccurate data but is pedagogically absurd and deeply disrespectful to young children's multiple expressions, competencies and possibilities. In Malaguzzi's terms, baseline assessment in whatever form has the tendency towards constraining, limiting and 'robbing' a young child's potentiality and their 'one hundred languages'. Baseline assessment and its image of a culturally 'poor' child tells a quasi-scientific, deficit-based story about the *in*competencie*s* of a young child (and by extension, their family) and potentially closes down who and what they might become.


NATIONAL
TARGET

Section 3: General proposals

This final section of part A presents some broad alternatives to the current system. These short articles present a critique of primary school tests, reconsider the educational purpose of assessment, and suggest, across the full primary age range, what might better fulfil this purpose.

The section begins with the position statement announced by *More Than A Score*, a broad coalition of the largest teacher union the NUT (now part of the NEU), curriculum associations, campaign groups and academic researchers. This was published early in 2017 as a challenge to Government to rethink the aims and modes of assessment and its link to accountability.

Around the same time, reacting to public dissatisfaction including parents' protests, the Education Select Committee of the House of Commons launched an enquiry. It invited not only comments on the existing tests but suggestions for improvement. The second text in this section is an extract from one of the hundreds of responses. After some specific criticisms of the 2016 tests, suggestions are made as to how assessment could encourage a broad, interesting and age-appropriate curriculum.

More specifically, John Richmond has compiled a set of alternative proposals for the assessment of primary language and literacy. Extracts are published here.

In contrast to English practice, the final article describes testing in Denmark and the protections against it becoming high-stakes. Its computer-based 'adaptive tests' adjust to an appropriate level for each student. (Interestingly, the author is currently providing consultancy advice for assessment reform in Wales.)

Assessment – what we stand for

More Than A Score

England's system of statutory primary assessment has been in place for 25 years. Most teachers and most parents have known nothing else, and it has become difficult for us to imagine a different way of doing things.

- Can we really do without our SATs?
- Wouldn't standards fall if schools weren't held to account for their results?
- Wouldn't teachers slacken and pupils regress?
- What would managers and policy-makers do without the information that tests provide?

More than a Score says: yes, we can manage without SATs, and the whole battery of other assessment instruments that dominate the life of schools.

The current system of testing every individual child in order to judge the effectiveness of teachers and schools is deeply flawed, and has had negative effects.

It focuses the energies of pupils, teachers and parents on achieving success in limited aspects of a narrow range of subjects: the school curriculum is dominated by Maths and English, and these subjects are themselves distorted by the need to make them testable. When schools are judged primarily on statutory test results, pressure and stress builds up on pupils and teachers alike: the system becomes punitive.

To leave this system behind would not be a leap in the dark. Other countries do things differently. We can learn from them.

Across the world, and in England too, there are educationalists who have thought deeply about assessment issues, and concluded that there are better

ways of achieving excellent teaching and learning than our present system allows.

What should assessment do?

- We want assessment that supports children in their learning – and enables teachers to identify pupils' attainment and learning needs.
- We want assessment that treats young people in the round as whole persons.
- We want modes of assessment that are appropriate to children's development.
- We want assessment which helps to identify schools which need extra support.
- We want assessment that enables a dialogue between parents and teachers.
- We want assessment that enables schools to develop improvement strategies in line with their own values.
- We want assessment that tells us about national standards of attainment across the whole curriculum.

What does the system need to look like to achieve this?

No one test can reasonably perform all these tasks. We need different forms of assessment for different purposes.

In the classroom, we want to see both formative and summative assessment. Formative assessment is ongoing assessment that supports pupils while they are learning. It is based on observing what children can do, and discussion and feedback between learner and teacher. Summative assessment tests pupils to find what they have learned at a particular point in time – at the end of a project or unit of work, for instance. Teachers should be trusted to use their professional expertise in determining the best methods of assessment. In some countries, summative tests can be based on national 'question banks'.

Formative and summative assessments can be combined in an approach that is detailed, rigorous and supportive.

We want an assessment system which enables teachers in different schools to compare the progress made by their pupils, against national standards. This can be done by teachers coming together to moderate pupils' work. The results of moderation will feed into a school's self-evaluation and plan for self-improvement. This in turn will be assisted by supportive inspection of schools.

Parents should be acknowledged as partners in children's learning and need information that enables them to support their children's learning. For reports to be meaningful to parents, they need to summarise what children can do and understand. Some schools already aim to produce rich, detailed descriptive reports on pupils' progress, that use the outcomes of formative and summative assessment to inform feedback to parents and pupils, and to plan learning development. Assessment in the early years, culminating in the Early Years Foundation Stage Profile, offers an example of an approach that can be used to track children's learning throughout primary education.

We propose that national monitoring of standards should involve testing only a sample of children; when it comes to the evaluation of national standards in Primary Science, this is what the DfE already does! There is a need to monitor the standards of the primary school system. But there is no need to impose highstakes testing of every child to provide this information. Tests could include different curriculum areas so that a picture of standards across the whole curriculum would become available, informing teachers' work.

Where would these changes take us?

With changes like these, we would have an assessment system which covered the whole range of children's learning, not just a small number of core subjects.

- It would be a system that encouraged teachers to think inventively about children's learning and how to support it.
- It would avoid the negative impact of high-stakes testing on children's mental health.
- It would form part of arrangements for a different kind of accountability, which combined school self-evaluation with ways of reporting to stakeholders outside the school.

The voices of those calling for changes like these are growing louder and more various. Most of the business world wants learners who are well-rounded and creative. Parents are increasingly frustrated by the way the school system works against their children's development. Even within the current system, many teachers are working on practical alternatives to the testing culture.

For too long, the needs of external testing have dominated pupils' entire experience of school. Successive governments have failed to ask a crucial question: what kinds of assessment create the conditions for young people to thrive in an uncertain and innovation-rich world? It is time for our energies to shape an education system in which such a question can be answered.

More Than a Score calls on the Government for an urgent, thorough review of assessment and accountability of primary schools.

Some modest proposals

Terry Wrigley

Primary school assessment has been brought to a point of crisis by a combination of a poorly revised *National Curriculum* and the 'high stakes' ways in which testing is locked into a *wider system of control*. Both have exacerbated the damage to children and education, and encouraged 'teaching to the test' and curriculum narrowing.

The consequence of a curriculum clumsily planned to make England a 'global winner' in PISA, and the failure to consider children's development, is a high failure rate which is seriously *demoralising* as well as test results distorted by each child's age.

Specific problems with current test design include:

- a phonics check which *relates poorly to real literacy*;
- a KS2 reading test which is *remote* from children's experience and biased against disadvantaged groups;
- spelling and grammar tests which *relate inadequately to children's writing*;
- and writing assessments which encourage *formulaic writing* rather than high quality communication.

Assessment needs to be made *fit for purpose*, with an emphasis on assisting children's individual *progress* and fruitful communication with their parents. *Trust and professionalism* need to be encouraged.

Reading KS2 (2016)

Passage 1 was based around two children straying from a garden party in the house formerly owned by the girl's family, and rowing a boat across a lake to find a statue commemorating her ancestor. Such a situation would be

inconceivable to many children - a garden party, a garden large enough to contain a lake, owning a house, an ancestor, a statue of someone in your family?

Many sentences set up an ironic distance for the reader from a more normal situation or meaning, which requires the pupil to be familiar with that expected norm. Consider the following (passage 2): 'She adored warthogs but their Hollywood movie star eyelashes didn't fool her.'

Some sentences depend on children already having substantial cultural capital, probably acquired outside school e.g. 'many of the artists had no knowledge of *natural history*', 'Mauritius… was *spat* out of the ocean floor by an *underwater volcano'*.

KS2 grammar

Most children use a wide range of clauses fluently without being able to name them as coordinated or subordinated, temporal or concessionary. Similarly, almost all children use modal verbs in various tenses appropriately and sensitively before starting school. They don't need a grammar test to learn this.

Conversely, children are being drilled on issues that 11-year-olds are unlikely to use. Almost the only subjunctive they would use would be the formulaic "If I were you..."

Writing KS2.

Quite appropriately, this assessment was carried out by teachers and based on a portfolio of work. It is ironic, therefore, that the Standards and Testing Agency insisted on imposing tight criteria. This resulted in classes spending months re-editing writing to match them, rather than improving ideas and

expression. Teachers rightly complain about 'shoehorning' fronted adverbials, subjunctives and semi-colons into texts to be able to award higher levels, resulting in dull and formulaic writing. The deep problem is a lack of trust in the teaching profession.

Some realistic proposals

A) The DfE should work with the teaching profession and other experts to establish new structures of quality control which are more positive and less punitive. The emphasis should be on improving professional judgement rather than external top-down control.

B) It is important to have shared standards, but these should be expressed in terms which are aspirational. As an example, Finland's national curriculum describes '*good performance'* at key points, rather than 'meeting / not meeting expected standards'.

C) The key aims of assessment should be discussed with the teaching profession, and the various purposes distinguished. For example, in order to evaluate overall standards and improvement nationally, sampling would be sufficient; it could be more detailed without overburdening individual children and would avoid distortions due to 'teaching to the test'. On the other hand, diagnostic assessment should not result in numerical scores which lose the key information. Teachers should have access to a bank of assessment tools to complement or verify their own ongoing observations.

More specifically:

D) The phonics check should be abandoned, as it is too narrow and gives poor data and information. Year 1 teachers should be expected to carry out diagnostic assessments of various aspects of reading (phonics, irregular word recognition, breadth of vocabulary, expressiveness in reading aloud, attitudes to reading) using their own observations, flexible tools such as miscue analysis, and, if necessary, some test items drawn from a national assessment bank. Information should be shared with parents descriptively, not as pass or fail.

E) Assessment of writing should be based on authentic purposes. Rather than mismeasuring through lists of fixed artificial criteria, teachers' evaluations should be strengthened through guided moderation involving training, local panels to review sample scripts and visiting moderators.

F) Separate tests of grammar, punctuation and spelling should be abandoned, and teachers expected to draw from banks of test items as they see fit to supplement their assessment of writing.

G) The assessment of reading by the end of Year 6 should reflect a wider range of genres and purposes than at present, including more extended texts, reading for information (locating, selecting, modelling etc.) and critical literacy, as well as non-print media. It should be based primarily on teacher assessment underpinned by moderation (see E above) and focusing on authentic reading activities, and with a bank of test items for optional use. If designed to reflect different standards of achievement, these might be judged as met at an earlier stage rather than all at the end of Year 6.

H) A challenge should be set to pupils in various school years requiring knowledge and skills from several subjects and drawing on elements of literacy / communication / mathematics. The products and processes would indicate progression and development over time, provide feedback to the child, and facilitate discussion with parents and with the next year's teacher or school. This would create a more balanced assessment and provide opportunities to assess critical and creative aspects of learning.

I) Teachers should collect samples of work for each child across the curriculum, to be passed on to Y7 teachers in secondary schools. Like H above, this would avoid curriculum narrowing.

J) Assessment should serve teaching, not the reverse. None of the above procedures should be made so elaborate that it places a strain on teachers' workload and distracts from teaching.

Assessment in English 3 to 11

John Richmond

General principles

Curriculum and assessment have an interactive and mutual influence on one another. A central principle ought to be: decide on your curriculum first; then decide how to assess progress within that curriculum effectively. Too often, the order of priority of attention to the two things has been the opposite.

I offer here a critique of current arrangements in primary assessment, followed by practical, educationally preferable alternatives.

Early Years Foundation Stage

One tool for formal assessment at this stage is the EYFS Profile, which has been in operation in one form or another since 2003. It accumulates findings about a child's achievements throughout the reception year.

The profile is a broadly enlightened instrument, based on admirable principles:

- Assessment is based primarily on the practitioner's knowledge of the pupil. Knowledge is gained predominantly from observation and interaction in a range of daily activities and events.
- Responsible pedagogy must be in place so that the provision enables each pupil to demonstrate their learning and development fully.
- Embedded learning is identified by assessing what a pupil can do consistently and independently in a range of everyday situations.
- An effective assessment presents a holistic view of a pupil's learning and development.

- Accurate assessments take account of contributions from a range of perspectives including the pupil, their parents and other relevant adults. (Early Years Foundation Stage Profile: 2016 handbook, p7)

A section of the guidance entitled 'Responsible pedagogy' contains an eloquent statement of the right relationship between teaching and assessment:

> Responsible pedagogy enables each pupil to demonstrate learning in the fullest sense. It depends on the use of assessment information to plan relevant and motivating learning experiences for each pupil. Effective assessment can only take place when children have the opportunity to demonstrate their understanding, learning and development in a range of contexts.
>
> Pupils must have access to a rich learning environment which provides them with the opportunities and conditions in which to flourish in all aspects of their development. The learning environment should provide balance across the areas of learning. Integral to this is an ethos which
>
> - respects each child as an individual
> - values pupils' efforts, interests and purposes as instrumental to successful learning (p8).

In each of 17 'early learning goals descriptors', teachers are required to judge, at the end of a child's reception year, whether he or she is meeting or exceeding the expected level, or is best described as being at an 'emerging level'.

The goals are grouped within three 'prime areas of learning':

- communication and language
- physical development
- personal, social and emotional development

and four 'specific areas of learning':

- literacy
- mathematics
- understanding the world
- expressive arts and design.

These are combined with 'a short narrative describing the pupil's three characteristics of effective learning'

- playing and exploring
- active learning
- creating and thinking critically.

There is a discussion to be had about whether this is over-complex, and I would favour a simplification.

The overall excellence of the intention of the profile is spoiled, so far as the judgements on literacy are concerned, by the intrusion into the goals for reading and writing of the government's overriding obsession with phonics. These judgements should represent a broader understanding of how young children's powers of literacy develop. However, to stick to the bigger picture, the Early Years Foundation Stage profile gives Key Stage 1 teachers ample information as to the achievements and needs of their new pupils.

Key Stages 1 and 2

The government's requirements and plans from summer 2016

There is no space here to comment on the shifting recent history of the government's requirements for assessment at these Key Stages. I'll confine myself to the requirements imposed from summer 2016. Unfortunately, these are no improvement on the past, and in some respects make matters worse.

Key Stage 1

Beginning in summer 2016, there have been tests for Year 2 pupils in reading and in grammar, punctuation and spelling. These tests are externally set, but marked by teachers in school. (The 2016 Year 2 grammar, punctuation and spelling test had to be abandoned, because it had already appeared online as a practice paper.)

Writing is assessed by teachers, with moderation, on the basis of pupils' work throughout the year.

The Year 1 phonics check continues, despite all its inadequacies both for successful and struggling readers.

Key Stage 2

Year 6 pupils take tests in reading and in grammar, spelling and punctuation. These are new tests, with only one version each (previously each had two versions), but including questions designed to test higher-achieving pupils. The tests are externally set and marked. Writing – understood as being somehow separate from grammar, punctuation and spelling – continues to be internally assessed, with 25% of pupils being moderated.

The grammar, punctuation and spelling tests divorce those three aspects of language from the contexts in which they should be considered: actual, whole, authentic texts, read or written.

Assessment of speaking and listening abandoned

The assessment of speaking and listening at both Key Stages, even in the unmoderated form which applied until 2015, has been abandoned completely.

An alternative proposal for end-of-Key-Stage assessment

Two tests at the end of each Key Stage

It is perfectly possible to test reading and writing, in all the aspects appropriate for a given age group, based around an appropriate selection of authentic texts and tasks. They would provide a balance between external and internal assessment.

I envisage two tests for each Key Stage: one for reading and one for writing. Teachers would select texts (for reading) and tasks (for writing) from a national online bank, updated regularly, to show the quality of their students' performance across a range of genres. The tests would be internally assessed, with moderation.

A reading test of this kind would assess pupils' overall understanding of, and response to, the meaning and structure of three texts in different genres, as well as their recognition of words, their understanding of grammatical concepts and terminology, their grasp of conventions of punctuation, and their apprehension of spelling patterns and families.

Similarly, the writing test would require pupils to write three pieces of continuous prose in different genres, with a suggested word limit for each. The writing would take place at any time within a half term, and could be blended into the wider curriculum. This would assess the extent of a writer's competence, not just as a communicator of meaning in different genres, but as a user of the conventions of punctuation and spelling, and as a controller of the grammar of English.

No need for separate tests on grammar, punctuation, spelling and phonics

The separate tests of grammar, punctuation and spelling and the Year 1 phonics check could then be abolished. Reading would be seen as what it is:

an activity in which the decoding of words and the comprehension of meanings are complementary, interactive aspects of the same, complex process. Writing would recover its wholeness too. Grammar, punctuation and spelling would be put back where they best belong: as integral parts of the construction of meaning in the written language by producers (in the writing tests) and by receivers (in the reading tests).

Performance descriptors linked to the alternative curriculum

Online performance descriptors of competence in reading and writing would be accompanied by examples. Their purpose would simply be to help teachers and moderators decide whether a pupil had *not yet achieved*, had *achieved* or had *exceeded* an expected standard in reading and writing. There would be two performance descriptors at each level, accompanied by examples, one for achieving and one for exceeding the expected standard.

Value talk as highly as reading and writing

Achievement in the spoken language is of equal importance to reading and writing. Recognising the difficulty of externally setting effective tests, I would supply schools with online performance descriptors of competence in the spoken language, supported by audio-visual examples. The descriptors and examples would help teachers and moderators decide whether a pupil had *not yet achieved,* had *achieved* or had *exceeded* an expected standard. Assessment of spoken language would be internal, with moderation, like that of reading and writing, but on the basis of pupils' achievements over the whole of Year 2 or Year 6.

In the longer term: trust teachers more

At some point in the future, once teachers have become familiar with these arrangements, the government might feel secure in relying on teachers'

professional judgements in making accurate assessments of their pupils' achievement without the compulsory use of externally set tests. The online banks of texts and tasks would remain, and be refreshed regularly, but it would be for schools and teachers to choose whether or not to use them. (The tasks might be helpful, for example, to newly qualified teachers teaching Year 2 or Year 6 pupils for the first time.) Whatever happens, local moderation will always be needed.

To conclude…

The *Early Years Foundation Stage Profile* is, overall, an excellent document, demonstrating an enlightened understanding of learning and of the relationship between learning and assessment. It is a little spoiled by the government's obsession with phonics as the only effective means of teaching early reading, and is perhaps over-complex, but no other assessment tool is needed at this stage.

At the end of Key Stages 1 and 2, the testing of reading and writing should treat these two complex activities as wholes. The current arrangements dismember them. Schools should assess their pupils using texts and tasks chosen from an online bank, updated regularly.

The spoken language should be assessed with the same rigour, using teachers' judgements of pupils' spoken language throughout Year 2 or Year 6.

The outcome should be judgements, moderated locally, as to whether a pupil *has not yet achieved*, *has achieved* or *has exceeded* an expected standard in reading, writing and the spoken language. Online performance descriptors, with examples, would help teachers and moderators to make their judgements.


FAIL!

National Tests in Denmark

Jakob Wandall

Introduction

Testing and test results can be used in different ways. They can be used for regulation and control, but can also be a pedagogic tool designed to refocus teaching, improve learning, and facilitate local pedagogical leadership. To serve these purposes the test has to be used low-stakes, and consequently the Danish National test results are made strictly confidential by law.

Danish is tested in Years 2, 4, 6 and 8; Maths in Years 3 and 6; English in Year 7; and Geography and Science in Year 8. (The later starting age means children in each Year are older than in England.) There are also voluntary tests in Years 4-8 to monitor progress in Danish as a second language.

Adaptive tests

Adaptive tests adjust to each pupil's level of proficiency during the test. In this design, the first item presented to the pupil has an average difficulty. If the answer is correct, the next item presented to the pupil will be more difficult. If the answer is wrong, the next item will be easier. In this way the test will adapt to the pupils' level, so that the sequence of items will be different for each pupil.

In a linear test where the series of items is predetermined, most pupils will experience that some items are too easy, others too difficult and some items that fit the difficulty of the individual pupil's ability. The items that are too easy or too difficult reveal very little.

Only the items whose difficulty fits the pupil's level contribute substantially to an estimation of his / her ability. In a well-designed adaptive test, most of the

items each pupil sees will be of a suitable level.

The Danish public school is a comprehensive school

The Danish public school system is built on the principle that pupils are not to be divided according to ability or social background. These divisions are prohibited by law. A consequence of this philosophy is that there is a considerable distance between top and bottom of the academic level in every class in practically every school. There is often a range of more than five years within each class.

How are the tests carried out?

The pupils have 45 minutes to answer as many items as possible. During this time, pupils typically answer 50-80 items. It is possible for the teacher to prolong the test for an individual pupil who is struggling.

Which parts of the subject are tested?

The tests are designed to assess large and important parts of the subject. However, not all aspects of a subject are suitable for formal testing, including pupils' ability to express themselves orally or in writing. Teachers are required by law to assess student progress regularly, so it is necessary to combine different kinds of assessment and evaluation. The national tests can only cover a very small part of the total need for evaluation in the Folkeskole. The website www.evaluering.uvm.dk also contains a description and a user guide to a large number of other evaluation tools.

Every subject is divided into 3 dimensions or strands, called profile areas, to make a more detailed and precise evaluation of the pupil's proficiency possible. These strands are mixed together randomly in a test. This means, in effect, that three separate adaptive test sessions are conducted simultaneously, but the selection of the next item in a given profile area depends solely on the

pupil's responses in the same profile area.

What is at stake?

In high stake testing, security, equal terms and fair conditions are key issues. But if the main purpose – as in Denmark – is to assess student proficiency in order to focus teaching, and thereby improve the conditions for learning, the teacher should have control over the test conditions (e.g. which aids, tools, remedies and assistive technology are allowed).

Indeed, if it improves analysis of student proficiency, it makes sense that the teacher is allowed to help, for example if the pupil gets stuck – even though it could have significant impact on the test score. Under these conditions high stakes uses could not be supported because the scores are not necessarily comparable.

When only a small part of the curriculum is tested – as in the Danish national test system – it is important to keep the stakes low. High stakes would result in too much focus on the tested profile areas and too little focus on creative, innovative and oral skills - which play a significant role in the curriculum of the Folkeskole.

Disseminating results

Only those who need information about the results for professional reasons are allowed to see them. All the results are kept in a secured database. The database contains all the items used for testing the pupils as well as the answers that the pupils gave.

Teachers have access to detailed reports with information about their individual pupils' results, as well as class-level results for their own class. The headteacher is allowed to see the pupil's overall results, the class results and

the results for the school. The local authorities have access to results at school level and test results aggregated to municipality level.

Parents must also be informed about their child's test results by the school. For this purpose, the computer generates a written description of the results for each pupil. There is a strong tradition of parental involvement in the Folkeskole, and the test results should support the school's cooperation with the pupils and their parents, in order to support each pupil in the best way possible.

What kind of response is given to whom?

The pupil: As soon as possible after the test (typically within a week), the teacher will talk to the pupil about the result. It is not just information but a discussion about what to do. Together they will plan the best way to improve teaching/learning in the future.

The most common situation is that the results match the teacher's perception of the pupil's proficiency; in this case, the test just confirms that they are on track. But experience shows that in a class, there are usually at least a couple of surprises, about which the test provides new knowledge and useful information.

The teacher: The results help teachers get a more precise overview of the impact of their teaching e.g. how are the pupils performing in that aspect of the subject which has been the focus of teaching? is there a need for any follow up activities? The test also provides information about pupils' proficiencies in areas to be taught in the immediate future, in order to match teaching towards the class's and the individual pupil's preconditions and needs.

The headteacher: The head has overall pedagogical responsibility at the school

and therefore an obligation to guide and coach individual teachers in pedagogical matters. The test results should therefore also be seen as a tool for pedagogic leadership.

National results: the national profile of performance

Assessment also helps monitor overall national development. Furthermore, schools and municipalities should be able to compare their results with the national average.

However, pupils' background in different schools is very different in socioeconomic terms, which has an influence on test results. This is taken into account and a statistical correction made. This correction considers factors such as gender, ethnicity, parent's education and socioeconomic status. The adjusted results are confidential but are shared with the school and municipality.

Recent experiences and plans for further development

Finally, it is important to emphasize that this system – like any other test system – provides information about the pupil's proficiency, knowledge and attainment but cannot provide ready-made solutions to pedagogical problems.

Part B Section 1
Formative assessment

The most important reason for assessment should be to stimulate improvement. First and foremost, it should give feedback to the learner. That is why formative assessment is often known as Assessment for Learning.

Although summative - and interim - marks and grades can also stimulate improvement, they give little sense of what the learner needs to do. All the valuable information disappears when it is condensed into a single number or grade.

Vague feedback is of little use. It can mislead learners into vague thoughts about working harder or avoiding mistakes. Praise is important but insufficient: the learner needs the kind of information which can be used to self-evaluate during future activities.

At its best, formative assessment is *empowering* for the learner. This is a particular emphasis in the chosen extracts. We hope readers will recognise some of the deep aims, and not only the familiar surface features such as wait time or traffic lights.

We begin with some examples from a small booklet by Paul Black and Christine Harrison *Science inside the black box*. It has a particular focus on science but also many useful general points which any teacher could use. The authors were part of a larger team engaged in exploring the principles and practicalities of 'Assessment for learning'. they had enormous influence, despite swimming against the tide of high-stakes testing. The project was based on collaborative work with teachers and shows practical ways in which formative assessment can open up the students' thinking processes.

Primary headteacher Flora Barton then writes about developments in her school to reduce the marking workload by focusing on verbal feedback, individually, in groups and to the whole class.

Finally, we present two short examples of democratic principles at work in Danish and German schools. They cast a new light on differentiation. In both cases, formative assessment helps students develop an understanding of their strengths and weaknesses, so that - with their teacher's guidance - they can choose their next task. Whereas high-stakes testing constrains learning, well developed formative assessment is expansive and liberating.

Science inside the black box: the work of Paul Black and Christine Harrison

Starting in 1998, Paul Black and other colleagues at King's College London were engaged in a major exploration of classroom assessment practices in order to encourage formative assessment. They worked in collaboration with 36 science, mathematics and English teachers in six secondary schools, leading to the publication of subject-specific advice in 2004 including the booklet *Science inside the black box* (still available for purchase from GL Assessment). This publication resulted from the understanding that while formative assessment has generic features, it also has some which are specific to a stage of learning or particular subjects.

Many of the generic principles are explained in another book from the project 'Assessment for learning: putting it into practice' (2003). For example, feedback must give clear prompts on how students can improve their learning. It is no use saying 'add more detail' if students cannot distinguish between relevant and irrelevant details.

The Science booklet is grounded in constructivist understandings of learning, recognising that teachers have to start from where the learners are, and that students need to actively reconstruct their ideas and not just accumulate additional facts. This means that the teacher must listen carefully to a range of responses from students "taking them all seriously whether they be right or wrong, to the point or zany, and helping students to talk through inconsistencies and to respond to challenges" (page 4).

To establish the contexts for useful feedback, Black and Harrison argue that learning experiences should be designed that challenge thinking and

discussion, involve rich questions, and encourage all the students to reveal their ideas. Discussion must be encouraged in both small group and whole-class situations. Some activities require students to apply scientific theories to new situations or see them from different perspectives. Others require them to predict outcomes and solve problems.

A range of suggestions are made, with clear illustrations provided of the type of question which will provoke thinking. For example:

i) Comparing (What is similar and what is different about combustion and respiration?)

ii) Categorising and recognising exceptions *(*Is is always true that green organisms photosynthesise?)

iii) Predicting (What might happen if this was growing in waterlogged soil?) (pages 5-6)

Although the writers acknowledge that teachers sometimes need to check the students' acquired knowledge, it is better to ask 'rich questions' that demand thinking and which require longer answers. For example

> Do you think friction is the same on the Moon as here on Earth? (p7)

Sometimes questions cannot be answered directly but require the learner to pursue a series of smaller questions and activities first. For example:

> Why are some forms of renewable energy more suited to some areas of the country than others? (p7)

The researchers asked teachers to experiment with wait time of around 3 to 5 seconds, and found that this led to longer answers, more students answering, and students commenting on or adding to other students' answers. It enriched thinking as it led to alternative examples and explanations being offered (p8).

Discussion in small groups provided a safe exploratory context for students to formulate and check their ideas first, before revealing them to the class. The researchers and teacher-partners experimented with cartoons to stimulate debate about alternative explanations. (Many interesting examples can be found by googling 'concept cartoons'.)

Students were encouraged to engaged with each others' explanations in dialogic ways, for example:

- Would Leon's method work in all cases?
- What sort of evidence would challenge Sally's idea? (p10)

Formative assessment operates in such learning environments in ways which are responsive to students' developing ideas, and sometimes it is hard to draw a dividing line between assessment and teaching.

Written feedback to promote thinking

Other parts of this research and development project related to feedback on written work. In their initial classroom observations, they had noted many examples of routine praise which did not engage with students' thinking (eg 'Good', 'Well done') or simple reprimands for poor layout (eg 'Date?' 'Rule off each piece of work'). They also discovered that giving a level or numerical mark did not generate improvements as students were unclear how they might improve. As the project proceeded, teachers became more aware of the importance of students 'taking ownership of their own learning' as well as becoming more aware of what they knew and didn't know.

Teachers' time was saved by getting students to self-check or peer-check basic tasks such as copying and labelling a diagram from a textbook. At other times, more writing was encouraged which would require students to articulate more

complex ideas. This would clearly require the teacher's professional response, but valuable suggestions are made which reduce the teacher's use of time whilst ensuring that the learners respond actively to the feedback. Various ways of providing feedback are suggested to provoke further thinking, for example *questions*:

> Can you suggest how the plant disperses its seeds? Could this be an advantage? (p12)

At other times, feedback suggested *where* students should *look for help* or how to improve. For example:

> Look back at the way we worked out scales for graph axes and pinpoint the mistake you are making. (p13)

Finally, based on this collaborative work between university-based researchers and teachers, the booklet suggests not overwhelming students with negative comments by using a 'two stars and a wish' approach. As an example:

> Circuit diagrams are clear and you can distinguish between a parallel and series circuit. Can you also explain why the bulbs are brighter in the parallel circuit? (p15)

Because formative assessment sometimes becomes ritualised into a set of tactics (eg traffic lights, no hands) when it is implemented hastily, *Science in side the black box* is a valuable resource as it contains a wealth of practical ideas which are grounded in principles of active learning and thoughtful exchange of ideas.

Verbal Feedback

Flora Barton

For feedback to be effective, it should be timely and provide children with their 'next step'. In our quest to refine processes whilst ensuring valuable feedback, research at our school has established the impact of verbal feedback to be far superior to that of any other. Verbal feedback can be immediate and not only provide children with their 'next steps' but more importantly ensure children understand *how* to achieve these.

We are not experts in assessment, but we continually adapt our practice to what suits our children and our staff. As educators it is vital to be able to reflect on what is working well and what isn't and then adjust practice as a result.

Purpose

During staff discussions about marking and feedback we have continued to contemplate purpose. Many schools have fallen into a habit of doing certain things purely for the benefit of Ofsted, or for senior school leaders. Our conversations centre on what our pupils need, what fits with the ethos of our school and what actually serves a *real purpose*. One of our core questions is: "What is the purpose of what we are doing and what are we trying to achieve?" This question helps us evaluate what we do, why we do it and the impact we are looking for. This means that we frequently challenge the status quo because there is often a difference between what is best for our pupils and our school to what is 'expected' elsewhere. When we discuss any policy change we not only carefully consider the impact on our pupils but also on teacher workload.

Teacher Workload

Marking and feedback has become one area where some schools have devised extensive written marking policies which they believe Ofsted expects. Teachers can spend hours every evening on just one set of class books. Marking is often the first thing teachers mention when they are asked what task takes most of their time.

The term 'work-life balance' is often mentioned in schools but many of the systems in place presently ensure that this 'balance' can never be a reality. Many have fallen into the trap of marking work not just to help children to move on, but to *prove* that children's work has been looked at. This is not effective feedback.

The senior KS2 teacher who first trialled verbal feedback is now coming to a full second year of not taking books home to mark. This one change has massively reduced workload. Further, the *impact* of consistent verbal feedback on pupil's progress has been phenomenal.

Evidence

During a whole class survey, 100% of the children agreed that a combination of verbal and written feedback (where they were noting down comments made during the verbal feedback session) was most useful, and that verbal feedback really helped them to understand exactly how they could improve and make progress. The children were able to explain that having the verbal feedback 'conversation' allowed them to ask questions to the teacher during their lesson time which immediately allowed them to improve their work or to fully understand what targets they were being asked to apply. They also said that having the time 'face to face' with the teacher put more pressure on them to focus on their improvements.

To evaluate impact, every pupil in this KS2 class was interviewed and asked to 'prove' their learning. They were able to talk through every book, explaining the feedback, their next steps, how they tracked their targets from piece to piece and how they applied the feedback given to their learning. Their progress could not be questioned, the impact and the purpose of what was being done was evident; the learning that was now being achieved within this one class was extraordinary. There are now many examples, across the entire school, of this increased level of pupil achievement.

Further, after having two consecutive KS2 external writing moderations, our assessment of writing was verified as consistent and fair. Both moderators made explicit reference to the feedback in the books noting the progress, depth and breadth of writing across all subjects.

Implementation

Any change in school can be very difficult to implement. Any changes to teaching practice can cause stress and strain on those involved, therefore implementation must be carefully thought through. It takes time for transformation to happen in schools and there are many aspects to consider; a transition time must be factored in for verbal feedback to become embedded.

A senior Key Stage 2 teacher was keen to pilot the use of verbal feedback. Having teachers willing to undertake research in their classrooms helps drive our purpose-seeking ethos. Therefore, making research a standard feature further helps when any change is introduced.

Our aim is for teachers to provide effective verbal feedback to all their children in maths and English once a week, either individually or through the use of feedback groups. Children are given timely and specific feedback, and teachers often have considerably less marking to do.

It is imperative that everyone follows a consistent format. The 'learning conversation' must focus on what needs improving. It should be a *dialogue* between teacher and pupil. Pupils need to fully understand what they are doing well (their successes) and what things they need to work on (their challenges). This is where the feedback loop is crucial and our overarching aim that "children must be clear about what they are doing well now, where they are aiming to get to and more crucially how they close the gap between the two" (Sadler, in Black and Wiliam 1998). Where feedback groups are used, children must be explicitly taught how to give and receive proper and useful feedback.

Finally, it is crucial that children are given the opportunity to apply the feedback they have received. The sooner they are able to implement their feedback, the more effective their improvements will be. It is about ensuring that children fully understand what is necessary to improve their work and how they can continue to make progress. As noted, for feedback to be most effective it must be timely, specific, clear, *purposeful* and focused.

Concluding Thoughts

A question that continues to be asked about verbal feedback is about when teachers are able to find the time to give it. The simple answer is that like anything, when you know something works and the impact is tangible, you find the time to implement it.

We continue to experiment with different ways of improving the effectiveness of feedback. In particular, methods are being explored, analysed and developed to maximise opportunities for different types of verbal feedback within each lesson and to ensure that it is factored into weekly plans as an integral and effective part of teaching and learning.

Involving learners in setting goals (Denmark)

Kirsten Krogh-Jespersen, Anne Birgitte Methling and Andreas Striib

Aims and evaluation hold a work process together, both for yourself and for your pupils. Just as you must formulate an aim for your teaching, so should each pupil formulate an aim which can be a leading thread in the pupil's work. This will form the basis for later self-evaluation.

At first, the children's aims are likely to derive from the teacher's and various pupils will formulate almost identical goals. However, the most important point, in the medium term, is that the individual learner or group of learners has ownership of the work process, knows what they have to find out and how to go about it. There is no way you can skip this, so you as teacher have to help them formulate their aims.

> A Year 2 pupil wrote and drew a book about owls. She was so enthusiastic about her book that she wanted to visit the nursery class and read them her book. She recognised that some parts of her book would need further explanation for the younger children. She was nervous, but had a great response from the nursery children.

An older pupil, who had just finished writing a story about mental illness, reflected on the process:

> Sometimes I had a few problems developing the action. I used to sit on a bench outside the classroom to think through the options. My aim was to get better at building an exciting plot. I acquired more understanding about mental illness through background reading. I read my story to my parents. I would definitely tackle something like this again.

Being positive about diversity (Bielefeld, Germany)

Annemarie von der Groeben

The Laboratory School is a comprehensive school for 5-16 year olds, based firmly on mixed ability teaching. Contrary to German norms, the school refuses to give grades or marks to children until this is required in the final year for transition to the next stage of education.

An option group studying Ethics is pursuing the issue of gender roles. Like the other secondary phase electives, it is open to pupils from four different school years. This tests differentiation strategies to the limit.

The first task elicits pupils' ideas and attitudes, specifically how they perceive the 'opposite sex'. Pupils choose from a menu of alternatives, or suggest their own. Every student should find their own personal way into the theme:

Task 1 (for everybody)

- The title might be: *She* (or *He*). You describe your dream partner.
- You can, so to speak, slip into the 'other skin' of the opposite sex, by imagining spending a day as a boy / girl.
- You could write a short essay on the theme 'If I had suddenly become a boy (girl), would I be a different person?'
- You can choose another way of dealing with this theme (but talk it through with your teacher first.)

Those pupils who are willing read their personal texts aloud to the group, while the others receive a confidential response from the teacher. Pupils are invited to give feedback to one another. We discuss what this means, what it is for, how to make sure it is experienced as helpful and constructive, what mistakes to avoid. Here are our agreed rules:

> Feedback must be helpful and constructive. It should give recognition and encouragement to the person who has just presented.

Step 1: check your understanding, say what particularly interests you ("I noticed that…"). No judgements at this stage please!

Step 2: give your personal impressions (When I was listening, I felt... What particularly impressed me was... I didn't understand...)

Step 3: Explain your judgements and recommendations (I thought you were particularly successful at... You could further develop... I think you should revise X... My advice would be...)

During this stage, the pupils are also asked to provide ideas for things they would like to study and ways of learning. These are collated by a team of four pupils who meet outside of the lesson to write up a common list. The trouble is, it would take years to get through them all, so a more manageable plan is agreed. This consists of some common activities for the whole class, and some choices for small groups to pursue in parallel with one another.

Whole class:

- Reflect about role models: which women and men are seen as role models by young people today? Why? (collective reflection)
- Expectations of the other sex: how do pupils imagine a partnership? what is the dream woman or man? (personal texts)
- Improvised drama: The first date
- Gender roles and education: how are gender roles established and promoted (using historical examples)?

Smaller groups:

- Happiness by order: analysis of contact ads in the internet and magazines (collect and interpret various examples, give a talk and lead a discussion with the whole class)
- Historic images of women: make an exhibition of pictures from various periods (develop a commentary and lead people through the exhibition)
- The emergence of the women's movement (give a talk)

- Controversies about gender: neurobiological and sociological modes of explanation.

We make available a collection of challenging texts, with alternative suggestions on how they might respond to them. These approaches enable the learners to engage with texts on a personal level, and bridge between abstract and experiential representations. We then move on to more theoretical approaches, looking at biological and cultural gender theories.

The course ends with an essay. The guidance not only helps students achieve quality, but also provides criteria for evaluation. For example:

> Is your reasoning clear, logical, convincing?
>
> Which of your own thoughts, experiences and questions are developed?
>
> What specialist knowledge have you brought in?
>
> Is the style suited to the theme?

Students receive an individual evaluation at the end of the year, addressed to them personally like a letter. It recognises their ideas, judgement and development during the course.

Summary

We have illustrated forms of differentiation which avoid ranking students or dividing them by 'ability', but which allow them to bring to their learning a variety of interests, talents, rior knowledge, research skills and viewpoints.

Formative assessment gives learners a sense of ownership of the aim, process of investigation and final product. It is expansive and motivates them to accept new challenges.

Whilst recognising diversity and offering different openings and pathways, it encourages all the learners to contribute to a shared understanding.

Section 2: Diagnostic assessment

Diagnostic assessment means the focused use of assessment to identify particular barriers in children's learning and suitable remedies. It is rarely used in a blanket way for all children, being more appropriate when particular children are encountering difficulties.

Ironically, the Phonics Check was introduced as a "light touch" diagnostic tool. Yet it has little diagnostic value. Indeed, the Department for Education have admitted that it is specifically designed to *promote* 'the teaching of systematic synthetic phonics', i.e. the teaching method favoured by Conservative ministers. This runs against established evidence: phonics is an important element in any literacy teaching but there is in fact no research evidence to suggest that *synthetic* phonics is more effective than any other kind.

It doesn't check whether children can recognise non-phonic words which are crucial in English (the, was, though), or that they can use context clues to decide on a pronunciation (eg low / cow). It says nothing about understanding or enjoyment.

It is irrelevant for the many children who are already reading fluently, and for those who aren't, it does not identify the causes of their difficulty. It does not even include common irregular words (even *the*, *is*, *was*, *any*) or words whose pronunciation cannot be predicted from the letters alone (eg *soup*, *out*, *could* and *soul* - each pronounced differently).

It has little relationship with real reading, for pleasure or for information. Although the percentage of children passing it has increased dramatically over the years, there has been no improvement in Reading results for the same set of children a year later at the end of KS1.

Far from being 'light touch', it is yet another high-stakes test. Ofsted draw on the statistical database to compare the school's pass rate with the national average, and the Schools Minister Nick Gibb, a fervent advocate, has 'named and shamed' low scoring authorities.

This section aims to explain the problems with the phonics check and examines more diagnostic alternatives.

Firstly, veteran literacy researcher Margaret M Clark scrutinises evidence on the Phonics Check, and on the synthetic phonics method to which it relates.

Then Jonathan Glazzard makes alternative proposals. He argues against its use with children who are already fluent, and for a more precise diagnostic approach with children who are struggling.

The third item presents an example of a genuinely diagnostic tool, known as *miscue analysis*. Our illustrations show how errors enable teachers to discover specific problems, whilst recognising that fluent readers sometimes substitute sensible alternatives that fit the context.

Diagnostic approaches are not by any means limited to literacy. The Bell Foundation has recently formulated a set of graded indicators of progression from one stage to another, which aligns with the DfE's existing Proficiency in English scales. We include illustrations of this important diagnostic resource.

Finally, it is important to recognise that there is no sharp boundary between diagnostic assessment and other forms of assessment. Diagnosis is integral to formative assessment. Observation of children's learning processes can provide essential information on what the student has understood and misunderstood. Diagnostic assessment has its specific tools but is also a particular attitude within other assessment activities.

Synthetic phonics and the phonics check: the evidence

Margaret M Clark

Over the years since 2006, synthetic phonics has become the required method of teaching reading in primary schools in England, and that to be emphasized by those training primary teachers. There are good reasons to support the value of phonics as part of teaching children to read, but no evidence that *synthetic* phonics is superior to other approaches.

Phonics instruction refers to literacy teaching approaches with a focus on the relationship between letters and sounds. Most children need some systematic teaching of phonics, but within a broad programme. The question here is whether phonics should be the only method employed in the early stages, the books on which the children are learning be confined to simple texts, and whether synthetic phonics instruction is superior to analytic phonics. The defining characteristics of synthetic phonics are sounding out and blending. Analytic phonics avoids sounding out, inferring sound-symbol relationships from sets of words.

Drawing on a wide range of research from the 1960s onwards I found little evidence for one best method of teaching reading for all children, and certainly not for the superiority of synthetic phonics as the method as opposed to analytic phonics. Concern has been expressed by many researchers about this approach, in particular with regard to learning written English with its complex 'deep orthography', in other words the many irregularities between sounds and spellings.

Since 2012 a phonics check has been administered to all children in year 1 (aged five and a half to six and a half years of age) and again in year 2 to any

who fail to reach the pass mark. A government-funded evaluation carried out by the National Foundation for Education Research (NFER) has revealed increasing emphasis by teachers on decoding non-words, but no impact on reading comprehension.

Thie pass mark of 32 has been made by increasing numbers of pupils, but this is probably because teachers are spending more time 'teaching to the test'. Despite the rising pass rate in Year 1, the pass rate in Reading SATs at Year 2 remained static. This supports the argument that real reading involves many more skills than simply pronouncing the letters.

There are many reasons to doubt the value of the Phonics Check.

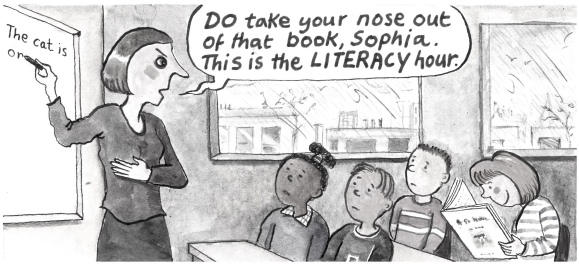
- Studies have shown that teachers are perfectly capable at identifying struggling readers without the test.
- The 'spike' in pupils scoring 32 (the lowest pass mark) compared with 31 (the highest fail mark) suggests that teachers are nudging pupils past the post due to obvious pressures. (In 2017 3% of children scored 28-31 marks, whereas 23% scored 32-35.)
- Pseudo words are known to disturb many children, including those who can already read well and also children with autism.
- There are striking differences between the pass rate of the oldest and youngest children in the class. In 2017 12% of September born children failed, compared with 26% of those born in August. Over three times as many of the youngest boys failed as the oldest girls. This suggests that many children are simply not mature enough for the test.

It is iniquitous that parents of five year olds are being told their children have 'failed' to read on the basis of such a test.

Despite reassurances that this is a 'light touch' diagnostic screening, school-level results are recorded on RAISEOnline and made available to Ofsted for use in inspections. Comparisons are made between the school's and national pass rates, even though the number of children taking the test in the school may be very small.

Recent research shows that the phonics check does not even test the full range of regular spellings, but only the most common. For example, even c pronounced as s (eg face) is missing.

The key purpose of the phonics check appears to be to enforce schools minister Nick Gibb's favourite way of teaching reading. There is no conceivable benefit from screening already fluent readers for decoding regular phonics - and their teachers know who they are. Conversely, children who are struggling may have various difficulties which need to be distinguished - a minimal grasp of the way print works, a struggle with common but irregular sight words, poor pronunciation of specific consonant blends (sp, st, etc.), among others. These children need genuine diagnostic assessment, not the simplistic and misleading phonics check.



What could replace the phonics screening check?

Jonathan Glazzard

The synthetic phonics approach introduces beginning readers to the smallest units of sound in words. Pupils learn to read words by enunciating each of the phonemes in sequence throughout a word to read the target word. This approach is often referred to as 'blending' or 'decoding'. Beginning readers are thus able to decode print regardless of whether they understand the words they are reading.

Such an emphasis on decoding is demonstrated through the introduction of the phonics screening 'check' for all pupils in Year 1 at the age of 5-6. This is a test of decoding rather than comprehension, and, in order to ensure that pupils are not reading words from memory, many of the words that are presented to children are pseudo 'non-words'. This means that the only way of identifying the target word is through enunciating the phonemes in sequence throughout the word and blending them together to identify the word.

Its imposition relates to the obsession of a government minister Nick Gibb for the synthetic phonics approach. It was designed to ensure that teachers all teach in an officially approved way. It is too inflexible to serve as a diagnostic tool to identify the learning needs of all young children.

The United Kingdom Literacy Association (UKLA 2012) recommended that the 'check' should only be used to identify development needs for individual children rather than being used with all children, on account of its holding back more able readers and potentially undermining their confidence as readers.

It does not take account of the different ways in which children acquire literacy. Although blending separate phonemes is a prime skill through which many children learn to read, many nevertheless learn to read through visual approaches and some use contextual cues as a basis for word recognition rather than relying on the skill of blending. This might involve missing a word out and reading ahead to the end of a sentence before going back to identify the target word. Some children are fluent in reading by the time they take the 'check', raising further questions about its relevance.

Its main feature is that it separates decoding from making sense of a text. This is a strength, if used on particular children at certain points in time, but it is also a major weakness.

The Simple View of Reading

The Simple View of Reading (SVR) proposes that reading ability or reading comprehension is the product of two components; decoding and language comprehension. They are independent of each other and each is necessary.

For children who are struggling readers, decoding is a better predictor of reading ability; but comprehension is a better predictor to explain variance in reading ability among skilled readers. For children who are already fluent readers, the check is inappropriate because they have already mastered the skill of word recognition.

As word recognition develops, there is a gradual shift from phonological processing to orthographic processing (for example, retrieving whole-word shapes from their store in the visual memory) and skilled readers tend to use orthographic strategies rather than phonological strategies which rely on grapheme-phoneme conversion. Fluency as well as accuracy is important in decoding, so this skill also needs to be developed.

Vocabulary is one of the most consistent predictors of reading comprehension: children with good vocabularies understand texts better, and the predictive relationship between vocabulary and reading comprehension increases through the primary grades.

Ehri's model

Ehri's theory of reading development concentrates on the initial difficulties of decoding, but does not look at language comprehension. It distinguishes four phases in the development of automatic word reading: pre-alphabetic; partial alphabetic; full alphabetic and consolidated alphabetic.

In the pre-alphabetic phase, children have not yet understood the relationship between phonemes and graphemes. At this phase their reading is dependent upon visual memory. They may be able to read environmental print, especially if it appears with salient visual cues such as logos which use specific colours and fonts.

In the partial alphabetic phase, beginning readers are able to identify the initial and final phonemes in spoken words and make some connections between graphemes and their corresponding phonemes. Their attempts at decoding are not always accurate at this phase but they are no longer arbitrary.

The phonics screening check fails to distinguish these phases. Thus, it fails to assess reading against a developmental framework. This does not help teachers to determine appropriate forms of intervention for children whose decoding skills are not secure.

Other diagnostic issues

In the case of poor readers with under-developed skills in decoding, assessing the skill of decoding (which is what the phonics screening check does) is

insufficient because skilled teachers will already be aware that the skill of blending has not been mastered. Other means are needed of assessing whether children have an appreciation of rhyme or alliteration, whether they can substitute different initial letters while leaving the rest of the word the same, and so on. Some of these judgements can best be made informally or in playful interactions with individual children.

Additionally, blending at the level of the phoneme - required in the phonics screening check - is an advanced skill. The check informs teachers whether children can or cannot do this but teachers will already know this through their ongoing formative assessments. Once we know that a child is unable to blend at the level of the phoneme, the Check does not help teachers to identify what they need to do next to support the child.

From a developmental perspective it is easier for children to process larger units of sound than smaller units. Phonemes are the smallest units of sound within a word. Instead, by adopting a developmental approach blending and segmenting at the level of the whole word is a logical place to start developing this skill. Children can be asked to blend and segment compound words (*tooth-brush/toothbrush*). Once this skill has been mastered they can progress onto blending and segmenting *syllables*. They can then progress to blending and segmenting at the level of the onset and rime (*c-at / d-og / s-it / c-oat*) before progressing to blending at the level of the phoneme. Teachers can use this developmental framework to assess what stage children are operating at within the skills of blending and segmenting and, more importantly, the stage of development informs them how to support the child. Teachers also need to assess the skills of phoneme addition, deletion and substitution as well as awareness of rhyme and alliteration.

Given the frequency of irregular words in the most basic sentences in English, immediate recognition of words such as *the, there, said, was*, and so on is very important. It is these words which hold sentences together, which carry the grammar without which meaningful literacy is impossible. Unless they are easily recognised on sight, partly by perceiving the shape of the word as a whole, fluency will be impeded. Assessing such key words cannot be done through a phonics check. If children do not have a good sight vocabulary teachers can adopt a developmental approach by assessing the sub-component skills which contribute to this. These skills include visual attention, visual discrimination, visual memory and visual sequential memory. If these skills are not secure then this could impede the development of sight vocabulary.

Teachers also need to assess children's linguistic knowledge. This is underpinned by vocabulary knowledge and sense of sentence structure. Children's reading development is influenced by exposure to spoken language and access to a rich language curriculum. Preparation for the phonics check becomes a major distraction for these children.

Conclusion

The phonics screening check is unhelpful in terms of informing intervention for the weakest readers and could have a detrimental impact on the progress of the most able readers who need to develop their reading *comprehension*.

For children who are working at the pre or partial alphabetic phases, a more detailed assessment tool may be required and should include skills such as compound word blending, syllable blending, onset and rime blending, phoneme addition, phoneme deletion and phoneme substitution. As reading is also a visual process, poorer readers should also be assessed against a

framework for visual skills development which includes visual attention, visual discrimination, visual memory and visual sequential memory.

For those children who can read aloud quite fluently, using a combination of phonic decoding and recognising frequent irregular words, the teacher's attention will need to be on a growing ability to make sense of texts in a variety of genres, as well as on the range of vocabulary. Assessment might be aided by informal questions about the meaning of texts and more difficult words.

Rather than constraining teachers through a single compulsory test, the 'phonics check', teachers need to be given professional autonomy to make choices from a battery of assessment tools. Skills in reading development vary across groups of children and individuals and the choice of assessment tool should be appropriate to the stage of reading development that the child has reached.

Miscue analysis

Lars was a big dragon (doggie). He was green and had red 11

eyes. He shot long (log) flames (flies) from his mouth (month). The grass 21

around (round) his cave was scorched (scratched). 26

Lars was the meanest dragon (doggie) in the land. He 35

scared (scratched) the people in the village (villain). At night the people 45

would look up to (at) Lar's cave. They saw the mighty 55

flames he breathed. He blew the smoke down to the 65

village. Often the people could not breathe. The 73

smoke was too thick. 77

Wind does many jobs that help us. It (skaters) scatters the seeds (for the) of plants in new (location) locations. It drives (wind mills) windmills and moves clouds, which are made up of water / vapor, to bring (s:snow) rain and snow.

Wind that moves very fast, however, is dangerous. (windstrohms) Windstorms can blow down buildings, (tir) tear up trees, and cause huge waves to crash on shores. The (wildest) wildest winds of all are (called) ^ tornadoes. A tornado is a spinning wind that (reach) reaches down to the ground. Scientist (s:sk:skeetist) believe that some tornadoes (revphlve) revolve at speeds of up to (three // three hundred) 300 miles an hour. These (wheeling) whirling winds tear and claw at everything they (taoch) touch.

There (size) are tornadoes of many sizes. Some (tepch) touch the ground (on) for only a few feet, but the average (tornadoes leave) tornado leaves a path one mile long and / fifty yards wide.

EAL Assessment Framework (Bell Foundation)

Primary school: speaking

Band B: Early acquisition / emerging

B1 Can answer yes/no questions (eg Are you hungry?) and 'choice' questions ('Do you want chicken or pasta?)

B2 Can produce simple joined-up utterances on known, familiar content or on topics related to personal opinions and experiences

B3 Can respond simply to a question relating to an immediate task, while syntax is basic and may contain errors (eg omission of verb inflection: 'She say he like Maths')

B4 Can repeat basic facts or statements previously learnt (eg reciting days of the week or answer a question like 'How many sides has a square?')

B5 Can deal with most day-to-day routines and common situations, and task-related language, where there is contextual support

B6 Is beginning to use forms (mostly first and third person present) of the verbs have, be, do, come, go and make, although not always accurately (eg 'I going play')

B7 Can give a short retelling of a story or sequence, perhaps fragmented, relying on objects and images, but may have difficulty with basic prepositions

B8 Attempts to follow and use simple modelled expressions in a small-group activity (eg 'You go first')

B9 Is beginning to meet the demands of group activities and class interactions without support for EAL (particularly when adults and role model pupils speak clearly and slow down their pace)

B10 Is beginning to participate independently in class discussions on familiar social and academic topics

NB The later items show the pupil is getting closer to the next band, but these are general indicators of progress and not expected to be achieved in a fixed order.

A complete copy can be found at www.bell-foundation.org.uk

Section 3 :
Supporting teachers in summative assessment

Summative assessment, as opposed to formative, means basically assessment at the end of a course to show how well the student has learnt. Although it often takes the shape of a written test or exam, this is not always the most valid method. (Imagine being asked to write an essay to prove you could swim.) Many skills can only be assessed practically, such as speaking a foreign language, baking a cake or playing the guitar.

This creates a problem because the Government mistrusts teachers, so practical teacher-led assessments have been marginalised or removed in formal qualifications in recent years. Thus, spoken English no longer counts towards an English GCSE grade.

If teachers are currently gaming the system, this is because the system fuels fierce inter-school competition. Quality assurance through 'naming and shaming' and using data to threaten schools is not conducive to balanced judgements. The transformation of assessment also requires a rethinking of the culture of fear and blame.

This section looks at various ways in which summative assessment by teachers can be supported. This is necessary in order to respect and accredit skills that written tests cannot deal with. We should note that schools in England are now out of step with assessment in other fields. Universities award degrees on the basis of a mixture of assessments, some of them written exams, some coursework or dissertations, some practical. The assessment is carried out by the university's own lecturers, and a sample is then checked by other lecturers inside the university followed by an External Examiner from another

university. Another example is the final qualification for doctors to become General Practitioners: a simulation, using actors, in which the candidate has to diagnose a set of illnesses and demonstrate how they would inform and advise the 'patient'. This is generally considered the best way to ensure that high standards are being met.

Two common ways of ensuring the reliability of assessment by teachers are through grade criteria and moderation. These serve to ensure fairness of judgement while facilitating different forms of assessment (oral, practical, etc.) as appropriate for the particular knowledge or skill.

Grade criteria are essential but they should not be treated dogmatically. The quality of the whole is more important than fulfilment of each separate point. It is ironic that the new SATs arrangements, from 2016, handed back responsibililty for assessing writing whilst imposing such stringent criteria that the writing purpose was undermined. Although the Department for Education intends to change this to a 'best fit', little has been changed as the student will still be expected to fulfil almost all of this unbalanced set of fixed criteria.

These criteria for Key Stage Writing mainly concerned technical features of presentation, whilst ideas and expression were marginalised. Redrafting took the form of shoehorning semicolons and fronted adverbials into children's writing, however inappropriately. The results were often formulaic and lacking in interest and creativity. This is well described in the first extract by Ros Wilson, a primary literacy consultant in the North of England, and reinforced in Nerida Spina's account of a parallel situation in Australia.

There are far better ways than this to support teachers in reaching reliable judgements. Criteria should be seen holistically, and related to valuable aims.

Moderation should serve as professional development, operating at different times and on different scales – before (i.e. using earlier samples) and after assessment, within and between schools. Developmental moderation is essential because criteria are rarely self-evident: a consensus needs to be reached on what they might look like in practice.

The third article in this section, by John Hodgson, looks back to a time when this approach was also common in schools. He explains the processes of teacher assessment which once extended to A-level English Literature. Moderation began well before the final assessment, as teams of teachers met with the Board's examiner to discuss sample scripts. Similar approaches, adapted for GCSE or primary school assessments, would quickly increase teacher professionalism, lead to intelligent and reliable judgements, and encourage more challenging and ambitious teaching and learning.

Within a high-stakes system, the demand for objectivity easily obscures curricular aims, narrows the curriculum and emphasises superficial or even trivial forms of achievement. The most recent example was the absurd ruling about the angle of the comma in a semicolon. (Of course, another source of the problem in the current school tests is the revised National Curriculum with its limited understanding of language and literacy.)

Our final two extracts illustrate well this problem of validity. John Hodgson criticises the use of separate grammar tests. John Richmond questions the assumption that grammatical complexity equates with writing quality by looking at the start of Dickens' *Our Mutual Friend* and *Bleak House*. How would Dickens have fared in today's school tests?

Assessment of Primary Writing in 2016

Ros Wilson

Of the 15 criteria for judgment that a child is 'working at the expected standard', 14 are pure Basic Skills (grammar, handwriting, spelling and punctuation). It would seem that the writing task was highjacked to provide a further Grammar Punctuation and Spelling test, rather than a fair measure of a child's creativity, originality, voice and style. The assessment process became a tick box exercise, with teachers playing 'Spot the full stop and give it a tick'.

Even the Department for Education's proposal to adopt a 'best fit' approach will not resolve this issue. It could mean that students still need to demonstrate 12 or 13 of the detailed presentation skills in their writing, without any greater emphasis on content, expression or style.

A very experienced writing consultant and former moderator who has also conducted informal research into the process this year recently mailed me:

> I fear the worst with the present process. I think back fondly to Natalie's 'Toad of Toad Hall' piece. That was a child behaving as a writer and a child with the confidence and ability to show her skills as a writer! Will the standards in the process produce writers like Natalie? I don't think so! Shame!

She was referring to the following piece of writing, in which the child has made a deliberate stylistic choice to print, despite her usual high standard of cursive script because, obviously, toads can't do cursive writing:

> 'I will surely perish in here, while this dark cloud of death hangs over me. Never have I realised how much I relished life on the riverbank. I miss the lush green fields, I miss the happy gurgle of the river, I miss the trips to your boat, I miss YOU, Mole and Badger, but most of all I miss the open road. I can feel the humming engines echoing past me in the world from above. Poop, poop...'

Does this score the 'Secure' judgement on the Interim Teacher Assessment Framework? No! Is it brilliant writing from a gifted eleven year old? Yes! The only criterion that comes anywhere near recognising Natalie's ability is the first:

- Creating atmosphere, and integrating dialogue to convey character and advance action.

And everyone who understands assessment knows that this is not an assessment criterion. It is, in fact, three criteria. And why is there only acknowledgement if the characterisation and action are conveyed through dialogue? Natalie is perfectly capable of doing that, but deemed that this piece was not the forum for direct speech.

I have spent much time in discussion with another writing consultant who has also assessed a large quantity of new writing samples. We came to the conclusion that there were three possible motives for the official criteria:

1. The authorities do not care about primary children becoming impassioned and talented writers – they just want them to be 'secondary ready' with perfect technical skills. Content, coherence and creativity are irrelevant.

2. The writing assessment is no longer an assessment of writing, it is a 'back up' Grammar Punctuation and Spelling test in case someone accidentally publishes the answers!

3. They have grown so fearful of the difficulties of assessing writing thoroughly and objectively that they only wish to assess those aspects of writing that work like most maths and science – right or wrong, tick or cross.

Teaching by numbers:
students' experiences of writing

Nerida Spina

The idea that their work was now coordinated by numbers rather than students and curriculum was worrying. Principals were mandating curriculum, pedagogy and assessment changes with a focus on explicit teaching of literacy and numeracy as in the tests.

Teachers confirmed that a great deal of curriculum, pedagogy, assessment and homework was now directed towards basic skills literacy. Time in class was being reallocated towards additional testing and data collection, particularly at the primary school. Outside the classroom, teachers reported spending extensive amounts of time – often on nights and weekends – recording and analyzing data, as well as engaging in new forms of work such as preparing for 'data conversations' with the school leadership team.

As an example, teachers made decisions to alter curriculum choices across a range of subjects to teach the required generic structure for 'persuasive writing'. A number of teachers referred to this work as 'doing persuasives'. In Australia, the ubiquitous 'PEEL (point, evidence, explanation, link) paragraph' was used by many teachers as the building block for teaching persuasive writing. This work began from Grade 2, when students are 7 years old. The emphasis on teaching the persuasive structure led one of the secondary school teachers to note that after 6 years of 'doing persuasives', students found it very difficult to write using any other generic structure. In her words, students are 'trying to persuade you every step of the way'.

The issue of teaching that was focused around the demands of standardised

assessment was raised by students themselves. A number of students said that creative opportunities were limited by the repetitive teaching of creative writing over the year. One said that:

> With narratives, sometimes we don't want to write them, but it's like . . . you have to write them. Which is good, because I guess it is an opportunity to show our creativity, but. . . it is like a disadvantage too. . . because we do them. . . every week. And then we have to write a whole new one! It's hard to be creative like that.

The students' responses were particularly striking given that the school has attempted to resist at least some of the pressure to standardise and limit curricular and pedagogic choices by including creativity and the arts as part of its strategic focus. Nevertheless, the focus on the technicalities of writing using a particular generic style meant that creativity was now difficult for students. As another student said, 'it [the technical approach to teaching creative writing] takes the fun out of it'. Another added: 'The thing is we have to do this every week, and it's like. . . dude can we just stop for a bit and focus on something else**!** ? Literally**!** We do it every single day**!** ' Although the students had been asked to comment on creativity, they continued to have a lively discussion about how assessment was indeed restructuring their school weeks. After a few minutes of conversation, one student said:

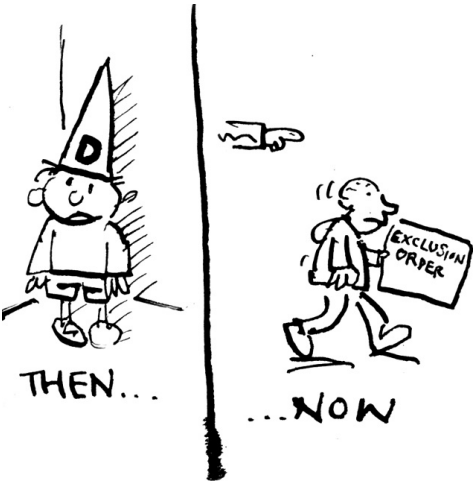
> I hate Mondays and Tuesdays because all we do is English and math. . . English and math. . . And like we have spelling tests on a Monday morning, but it's like, no wonder people don't do well on these pre-tests, because it's like, 'Get up, wake up and do tests, no matter how bad you're feeling.' It's a slap in the face. It's even if people are sick.

In a report to the Australian Senate's inquiry into the effectiveness of the

testing framework, an assistant principal similarly attested:

> Children sit many practice tests. In some classrooms they write each week in the test condition of 40 minutes with no assistance during this time.

The policies linking funding, performance management and data may place poorer performing schools at greater risk of being organised by numbers. Recent research has corroborated that creativity and self-expression are less likely to be present in schools with working- class children, 'while they flourish in schools where they are not present'.



Authenticity, validity and reliability in A level English Literature

John Hodgson

Introduction

This article comments on an A-level English Literature course that ran from 1977 to 1993. Over these years it embraced pedagogic and assessment practices intended to promote student reading and thinking rather than the rehearsal of received opinion. It was terminated by a Government move to reduce coursework assessment in national examinations.

To reflect at this point in history on the values and practices of a past A level English Literature course may seem a nostalgic indulgence; yet the course stays in the memory of many English teachers as a touchstone of quality as well as of innovation. It achieved a high degree of both validity and reliability.

The design and structure of the course and its assessment arrangements (which involved local moderators employed by the assessment organisation) gave students opportunities to show authentic personal response to their reading and capability in studying and writing a range of literary styles and genres. It provided little motive or opportunity for student malpractice, and engaged teachers regionally and nationally in a developed professional community of practice.

Origins and originality

AEB Syllabus 753 (as the course was originally known) started in 1977, for examination in 1979 (a two year course). The course was called "alternative" (the word was even printed on the exam papers), but it wasn't the first to introduce alternative elements. However, it was the first truly integrated alternative to other syllabuses. According to Bill Greenwell, whose tertiary

college in Exeter was one of the first to adopt the syllabus, this was largely due to Peter Buckroyd, the chief examiner, who had a vision of the whole enterprise. Not only did 753 offer coursework (initially a third of the assessment); it also had two open book papers, one of them also containing practical criticism (the Shakespeare paper). Students were allowed to bring annotated texts in to the examination.

Personal response to literature

Buckroyd grasped that open book exams needed appropriate questions. Rather than offer candidates merely a conventional question (such as a quotation of critical provenance, with the instruction "discuss"), Buckroyd's tasks drew attention to a debatable feature of the work and asked clearly and plainly, with prompts, for appropriate answers. Many questions directed candidates to particular pages, and suggested discourse features that the candidate "might like to consider". The tasks required candidates to look at detail and to support everything they said by textual reference. The "unseen" too became an invitation to detailed, considered response.

Coursework and consortium assessment

In addition to this "set text" study, candidates were required to write eight coursework essays of approximately 1000 words on books and tasks chosen in consultation with their teacher. The texts chosen had to cover all literary genres and to include non-fiction, and coursework tasks included opportunities for text transformation (such as pastiche and parody) and original writing. A further "extended essay" of 3000 words comprised a comparative study of two or more texts linked by author, period, or theme. By the late 1980s, 50% of the marks were allocated to coursework, which was guided and assessed by area moderators within country-wide consortia of A

level teachers. The venue was usually provided free by one of the centres, which also provided lunch.

The first meeting of the year, usually early in the spring term, was an in-service training event. This would focus on one or more aspects of the course, such as ways of approaching the chosen set texts, the choice of themes and texts for coursework study, organising and assessing students' oral presentations, and so on. Colleagues reported that these meetings were enjoyable because they offered opportunities for discussion of literature and ideas as well as of pedagogy and assessment. There was usually discussion of suitable coursework assignments, which were seen as a means of student learning as well as of assessment. Student and teacher were encouraged to negotiate the wording of a question, and the approach to be taken. Collaborative study could involve drafting and discussion with classmates and teacher, rather than an isolated effort to read the teacher's mind. The student's coursework folder was intended to be evidence of mastery of a range of texts and genres, including the student's original writing.

The second meeting of the year was given to moderation of candidates' coursework folders. These had been awarded provisional grades by the students' teachers. The majority of teachers attended these meetings, although they were not compelled to do so. (Schools and colleges could choose to accept the area moderator's assessment of student work without discussion.) Each school or college brought a sample of candidates' folders: these were discussed to indicate the standard to be held to in the succeeding internal moderation. Delegates to the meeting worked in groups, putting the folders in order of merit, and adjusting marks where necessary to ensure a reliable ranking across the consortium. In case of irresolvable dispute or

indecision, the area moderator would decide the rank order. Participants seemed to gain satisfaction from being part of the assessment as well as the teaching process, and the opportunity to see at first hand the work of their peers in other centres.

Creative writing and extended essay

Coursework offered opportunities to widen the range of students' writing as well as of their reading. Pastiche and parody, for example, allowed the writer to explore the text in a personal, affective way and to write in a creative mode, demonstrating grasp of form, character and theme. Creative assignments discussed in consortium meetings included the use of a minor character in one play as a major one in another (after *Rosencrantz and Guildenstern are Dead*); and additional dialogue occurring offstage, as in Harold Pinter's *The Caretaker*.

The most significant piece of coursework (in terms of allocated marks) was the "extended" essay, a comparative study of two or more texts normally linked by theme, period or author. The extended essay originally accounted for 11% and went on to account for 17% of the assessment. In Greenwell's view, this too was a brilliant innovation, for a variety of reasons. It gave the students a chance to do something approaching individual scholarship. Two, usually three books were studied, usually on a theme. The themes chosen might sometimes have been unexpected in an academic context, but, like the open-book exam questions, they gave students scope to develop their interests and responses. More than one adolescent horse-rider found interest in writing about *Black Beauty*, *St. Mawr*, and poems by Edwin Muir or Ted Hughes. Greenwell remembers more than one student who compared three novels about people with hearing impairment. Students had to read more widely, and so did the staff.

A few examples of extended essay topics chosen within local consortia include:

- fathers and daughters in Shakespeare
- an evaluation of Hardy's heroines
- self-determination in the face of oppression, as portrayed in the *Autobiography of Malcolm X*, *Beloved*, and the poems of Grace Nichols
- a study of children's readings of AA Milne and Lewis Carroll
- Stephen King's portrayal of small-town America in three novels
- and a study of banned and censored literature, focusing on *The Satanic Verses*, *A Clockwork Orange* and *Lady Chatterley's Lover*.

Students would prepare the work over a number of weeks, overseen by the teacher.

Validity and reliability

The course offered a high degree of validity and reliability in terms of educing candidates' responses to a wide range of literary texts and assessing these accurately. This was achieved firstly through open-book examinations that, in Greenwell's view, "killed question-spotting stone dead". The second guarantor of validity was the coursework component, which was woven into the course rather than single units of work for assessment. Eight of the nine pieces of coursework were simply essays produced as part of on-going classroom interactions: their validity derived in part from their not being specially worked up for the assessment. Students would normally write more than the required number of essays and would choose the best for their folder. Each essay was worth just over 4% of the total marks for the course, and thus was not a high-stakes assessment. For this reason, and because of the teacher's knowledge of the student's characteristic work, plagiarism and

cheating were rare. Moreover, as described above, the consortium system produced a developmental community of practice which ensured a high level of reliable, standardised assessment within and between centres.

Learning from the past

Twenty-three years later, we can see the widespread deleterious effects of governmental imposition of a narrow curriculum tied to a system of teacher accountability that breeds inauthentic practices. "Personal growth" is still the principle of English teaching that gains the allegiance of a majority of teachers, but the pressure on students and teachers to produce "results" ensures that the extrinsic value of a grade or mark matters more than the intrinsic value of authentic student creation and interpretation. The competitive, individualistic need to achieve a superior grade fuels a multiplicity of websites that will write essays for them for a fee.

In the current competitive, individualistic, inauthentic climate of what might be called institutionalised cheating, the AEB 753 (subsequently AQA660) English Literature A-level course stays in the memory of many English teachers as a touchstone of what validity and reliability might mean. It gave us control over at least part of the course. It enabled us to choose texts for and with the students, and to encourage students' authentic responses in a variety of genres. It enabled us to learn from colleagues while jointly discussing and assessing our students' work. Most of all, it gave a sense of personal purpose, allowing the talent and creativity of both students and teachers to be authentically validated.

Assessing primary literacy through grammar tests

John Hodgson

The Government's insistence on formal grammar teaching in primary schools has attracted much criticism. Many experts argue that the knowledge of formal grammar does not contribute to the development of language or literacy at this stage. Others claim that it can help, but only if taught in terms of expressive style in writing, not separately. Teachers are highly critical of many of the specific demands such as the ability to identify modal verbs, subjunctives or 'fronted adverbials'. This short extract summarises a few aspects of this critique.

There has been much comment, discussion and even fury in the media about the new grammar, punctuation and spelling (GPS) tests for primary school pupils. Parents, teachers, academics and other commentators claim that the tests are inappropriate for primary pupils and that these high-stakes assessments have a deleterious effect on teaching and learning.

Part of the problem lies in terminology. Children have to spot examples of grammatical constructions such as "fronted adverbials". This term has become notorious as it has not previously been used in grammatical descriptions and seems sometimes to apply to phrases that are essentially "adjectival". The deeper problem is that the label becomes more important than the underlying reality. It is obviously good to teach children the structures of language, particularly if such knowledge helps to express themselves more accurately. But testing a knowledge of labels is very different from testing an understanding of language structures.

Such understanding requires a connection between children's everyday understanding of language and the grammar they have to grasp. Linguists

such as Halliday have developed a functional approach to language that gives meaning to everyday interactions. However, GPS relies on 'ideal' forms of language that contradict everyday experience. The Oxford or 'serial' comma is outlawed when it is in fact common and correct usage. GPS requires that 'exclamations' must begin with 'How' or 'What' and include a finite verb - which is not the case in real language use. Terms like 'command' or 'exclamation', which have a social function, refer in GPS only to specific grammatical structures.

This context-free view of grammar implies that children's language is either right or wrong. GPS performance thus becomes a key indicator of a school's success or failure - even though the view of language enshrined in the tests is so limited.

Grammar and Great Literature

John Richmond

This extract from a speech by John Richmond at a conference of the National Association for the Teaching of English makes an important criticism of official assumptions about grammar teaching and language development. It illustrates how complex grammatical structures are not always the route to literary effect, and questions whether learning to write 'by the rules' is the best or only way to develop self expression and powerful use of English

.

> In these times of ours, though concerning the exact year there is no need to be precise, a boat, of dirty and disreputable appearance, with two figures in it, floated on the Thames, between Southwark Bridge, which is of iron, and London Bridge, which is of stone, as an autumn evening was closing in.

The opening of *Our Mutual Friend.* A magnificent sentence, full of subordination.

Equally effective as the opening to a novel is the following sentence:

> London. [full stop]

This is followed by another, even more effective sentence:

> Michaelmas term lately over, and the Lord Chancellor sitting in Lincoln's Inn Hall. [full stop]

and thus *Bleak House* begins to cast its spell. We read through its wonderful first paragraph, once again about London in the autumn, with ne'er a main verb to guide us. Second paragraph looms up:

> Fog everywhere. [full stop] Fog up the river, where it flows among green aits and meadows; fog down the river, where it rolls defiled among the tiers of shipping, and the waterside pollutions of a great (and dirty) city. Fog on the Essex marshes, fog on the Kentish heights.

Verbs here, yes, but only in the subordinate clauses. And so on for ten more lines (but only three more sentences); Dickens knew the use of the semi-colon. The third paragraph:

> Gas looming through the fog in divers places in the streets, much as the sun may, from the spongey fields, be seen to loom by husbandman and ploughboy. Most of the shops lighted two hours before their time – as the gas seems to know, for it has a haggard and unwilling look.

More verbs of various forms, but none main. It's too much of a flight of fancy, of course, but I'll mention it anyway, to suggest that there's any significance in the fact that Dickens only feels the need of a main verb when he introduces us, in the fourth paragraph, to the English law:

> The raw afternoon is [phew!] rawest, and the dense fog is [what a relief!] densest, and the muddy streets are [thank goodness!] muddiest, near the leaden-headed old obstruction, appropriate ornament for the threshold of a leaden-headed old corporation: Temple Bar. And hard by Temple Bar, in Lincoln's Inn Hall, at the very heart of the fog, sits [he does, at last!] the Lord High Chancellor in his High Court of Chancery.

When it comes to doing things legal, we need to get our language straight.

Section 4: Observation

Assessment through direct observation is well established, and can be used both for formative and summative purposes. Indeed, there is no real alternative to observation in many practical curriculum areas: consider, for example, sport or drama or speaking another language, where written tests lack validity.

In this section we explore various forms of observation, including video recordings but also spoken language.

Firstly, the Primary Language Record provides a well-developed approach to close observation of spoken and written language. It provides a far richer record of the learner's activity and development than numerical data. Rather than assessing just what is testable and measurable, as in the current National Curriculum assessment, it provides information that teachers can act upon. It also provides a framework for collaboration between teachers and parents.

Valerie Coultas then writes about the value of talk, rather than overreliance on written examination. She outlines ways in which spoken language can be assessed, but also suggests multiple opportunities for assessment in a range of subjects in primary and secondary schools. The extract concludes with a specific example of how to create an oral portfolio.

Finally Gawain Little and colleagues discuss the importance in mathematics of following students' thought processes, rather than simply marking their answers. They illustrate how superficial marking can provide misleading information on students' understanding. They then present three models of observational assessment which they have developed and used successfully in their school, including how to avoid excessive use of teachers' time.

The Primary Language Record revisited

The Primary Language Record was developed during the 1980s, and by 1988 had been adopted by several hundred Inner London schools as their main means of assessing language and literacy development. It continued to play an important role as the National Curriculum was introduced, and was originally the recommended model for National Curriculum English teacher assessment, as well as being taken up enthusiastically in the USA (alias 'Learning Record'). The Handbook sold over 100,000 copies.

Models such as this have been eclipsed by standardised assessment for accountability. National Curriculum assessment has moved relentlessly towards a narrow model where practising for the tests and raising scores dominates and where education has become the production of data to be analysed, mined and used for management purposes. Teaching to the test is now seen as good practice, and it is hard for schools to maintain a concern for the individual learner and for a broad curriculum with a place for the arts and wellbeing. The world of education is upside down.

The Primary Language Record provides a valuable model for rebuilding good practice. It is multidimensional, containing many ways of looking, and achieves its credibility by providing a full and nuanced picture of progress. Its feedback into the system is positive, it improves home-school relations, promotes pupils' learning and supports teachers' professional development.

Key principles

The Primary Language Record is a framework for recording ongoing observations of talking, reading and writing. It is formative, but also provides a cumulative record of progress.

The observations can also form the basis for a summative judgement at the end of a year or key stage. To assist this, illustrative criteria are provided for five levels of experience or fluency which encourage holistic evaluation. For example, descriptors for the highest level of writing from years 3-6 include:

- an enthusiastic writer who has a recognizable voice and who uses writing as a tool for thinking
- making conscious decisions about appropriate forms and styles, drawing on wide experience of reading
- able to craft texts with the reader in mind and reflect critically on own writing
- using mostly standard spelling
- managing extended texts using organisational structures such as paragraphing and headings.

The emphasis is on positive recording of what children *can* do, their enjoyment of work, and recommended support and next steps. Errors are viewed as information, enabling teachers to understand children's needs.

Holistic understandings of development are encouraged by drawing teachers' attention to five dimensions or strands:

- Confidence and independence
- Experience
- Skills and strategies
- Knowledge and understanding
- Reflection.

These cannot be separated out: for example, confidence depends on deploying appropriate skills and drawing on experience. By contrast, the National Curriculum tends to isolate Knowledge and Understanding,

neglecting the importance of the other factors as a foundation for successful learning.

Language and literacy development is viewed as situated (eg 'how the writing arose'). Observation is focused on natural behaviour in favourable and familiar contexts, to see what children *can* do when they have the opportunity to do their best. By contrast, high stakes assessment involves a narrow range of children's learning tested out of context, involving tasks designed so that many will fail.

Evidence is collected in inclusive ways, including:

a) evidence from homes, through two discussions with parents each year - making a big difference to home-school relationships and to teachers' knowledge of children

b) evidence from children, through two formal 'conferences' a year but also more spontaneous reflections during the year

c) information gathered about children's home languages and literacies

d) sensitive and appreciative recording of even slight gains made by children with special needs.

The format of observation sheets is flexible and open, but with helpful prompts to sustain the key principles. For example, for Speaking and Listening, teachers are encouraged to include different kinds of talk (planning an event, expressing feelings, telling a story etc) and to be aware of social dimensions (eg initiating a discussion, encouraging others, qualifying your initial ideas).

Problems and benefits

The main problems encountered have related to workload, and keeping descriptive records on every pupil can be hard to manage in the classroom. Various ways of overcoming this difficulty have been tried, such as keeping

detailed records for a third of the class each term, while maintaining a more general record for the rest of the class. Other possibilities have been to focus on a particular aspect of children's learning in different terms or, with older students, to rely on students' self-assessments in the areas covered by the record.

However, teachers have found it time well spent because of the rich evidence it provides them with, on which they can base their teaching both of individuals and of the class as a whole. The reflection related to this kind of record-keeping is also a highly effective form of professional development. It is a much better and more satisfying use of time than the endless data-gathering demanded of teachers in a performance-driven system.

An example of the record sheets can be downloaded at https://www.tes.com/teaching-resource/the-primary-language-record-11113280)

Assessment through talk

Valerie Coultas

Written exams have become the dominant mode of testing and other modes of assessment are being neglected. Re-establishing the importance of assessment through talk would promote good teaching and empower both pupils and teachers.

Oral assessment has been used in schools for many years. In nearly every lesson a teacher uses questions, at some point, to establish whether or not the pupils understand the topic or concept being taught. Drama teachers evaluate role play, improvisation and performance. Modern Languages has separate attainment targets for speaking and listening. English assessed speaking and listening at GCSE until recently, when a political decision was taken that the oral grade would no longer contribute to the final GCSE English grade.

Until recently, the English Literature AQA GCSE also assessed pupils understanding of literature through talk. The oral response option allowed the teacher to interrogate the pupils closely to ensure they had studied the text at a deep enough level to be awarded a particular grade. Through a presentation or a discussion, the pupil had to show insight into dramatic action, characters, setting, context or themes. Media Studies GCSE also has practical assignments that can include assessment through talk, for example if the pupils are asked to simulate a news team to create a radio news programme. When working as a Head of English, I found that pupils enjoyed the challenge of these oral assignments and felt an immediate sense of accomplishment.

Before the National Curriculum there was much greater flexibility in the use of oral assessment. The CSE mode 3 and the Certificate of Extended

Education were qualifications designed by teachers, with more opportunity to include oral assessment modules in a range of subjects.

The Cox Report (1989), which informed the first version of English for the National Curriculum, did not in fact advocate the kind of rigid written SATs and exams that have since been imposed on children and the teaching profession. It suggested that teachers should choose from a bank of SATs covering the three attainment targets. In primary schools, it was expected that pupils' responses should be mainly oral or practical unless the target required some writing or graphical work. The original report suggested that the task should be conducted over an extended period, and should reinforce teaching and learning and not be a bolt-on activity. The committee also suggested that coursework should have a major input into assessment.

What a difference between this and the present testing regime? Why did we move from some reasonably sane educational ideas to the dreadful, dreary SATS and exam papers? How come speaking and listening has once again become the Cinderella strand and been downgraded in GCSE English?

The reason is very simple. Speaking and listening and assessing reading aloud have to rely on teachers' judgements and Conservative prejudice will not allow this. Speaking and Listening is the educational casualty of the drive towards tight centralised control. If you want to use assessment to create crude league tables and 'name and shame', you have to have standardised written papers. The political imperative drives the agenda, not the needs of the pupils or good teaching.

In the past, teachers would listen to seven-year-olds read aloud to make a judgement about their decoding skills, fluency and comprehension. This is far more accurate and revealing than a phonics test of single words.

At 11 and 14 it would be quite possible to develop an assessment based on speaking and listening, drama or group work that incorporated reading, response to literature and writing. The teacher might challenge the students, after actively studying a text, to write in role as a character. This could be prepared by a speaking and listening activity such as hot seating. This would assess all the language modes in a single coherent activity, and would teach the pupils to plan and produce their best writing. The assessment is integral to the teaching.

Such methods would tell the teacher a lot more about the pupil's potential and give accurate feedback on how to improve. Such assessments stimulate collaborative thinking and encourage originality, evaluation and problem solving. These higher order skills are valued in the workplace and will help pupils to enter the adult world with more social and academic confidence.

There is no reason why every subject could not adopt an oral component as part of the system of assessment. Why not get the pupils to demonstrate their ICT skills by presenting a topic to the rest of the group? Why not arrange a debate on votes for women, with pupils in role as Nineteenth Century politicians? Why cannot a painting be researched, analysed and introduced to the class by a group of students rather than the art teacher? Such activities can create memorable learning moments for students. They learn more by finding out and teaching others than by absorbing information. Students will listen closely to their peers, particularly when they know that a lot of preparation has taken place beforehand.

Let's start thinking out of the exams box and use our knowledge of what really constitutes good teaching and learning to create wider and more

developmental forms of assessment. Can't we get the pupils talking about what they know rather than always having to write it down?

Example: creating an oral portfolio

1) Planning

Add a new character in a favourite TV series. First view the introduction to the programme together and discuss one of the following: colour; camera; character; sound or story.

(This activity could be organised as a jigsaw with each group reporting back on their topic. It would encourage reader response, build on prior knowledge, help to fill in knowledge gaps and promote whole class discussion.)

Next work in pairs to create a new character, decide exactly when and how that character would arrive in the sitcom and write a short script or storyboard of the moment when they first arrive. You can draw sketches of your new character and describe them in a commentary.

(The teacher could tell them that a TV script writer was coming to judge their ideas and ask them to prepare a formal presentation. A real scriptwriter could be invited, or the class teacher or another teacher could arrive in role. The teacher could prepare the students for the presentation by discussing formal and informal language and the different registers we adopt for different situations. The teacher's role is to provide all the students with the opportunity to succeed on such an occasion.)

2) Observing

Your teacher will observe some groups discussing and planning.

(The teacher should focus on one or two groups, while another group records its conversation in another room.)

3) Recording

Each group in turn will present their character to the whole class.

(The teacher has already begun to record the exploratory talk but now she needs to record the presentational talk at the final session.)

4) Pupil reflection and evaluation

You should now record your self-evaluation in a talk diary or oral portfolio. Think about

how well you engaged the audience, your use of language, the effectiveness of your body language, tone of voice, pace of delivery and where you might want to improve.

(This could happen in groups or individually. Evaluations can also reflect reading and writing.)

5) Making judgements and reporting

Your teacher will need a record of your activity to check her judgement with other colleagues, including moderators from other schools.

(The teacher has the evidence to support judgements. These will be more reliable if oral portfolios include a range of tasks and video recordings can be compared among colleagues.)

Maths is more than getting the right answer: redressing the balance through observation

Gawain Little, Jo Horn and Steph Gilroy-Lowe

High-stakes standardised tests are distorting primary school mathematics and failing to promote pupils' cognitive development. Greater use should be made of observation of process and product, for both formative and summative assessment, in order to recover an emphasis on reflective mathematical understanding and problem-solving.

First, we would like to start with a personal anecdote from one of the authors.

> I love maths and have spent the last few years teaching maths in a primary school, developing a new maths curriculum for the school, and coaching teachers in effective classroom practice. But I haven't always loved maths.
>
> At primary school, I struggled with times tables. Rote learning did not work for me and no matter how many times I repeated my tables, they wouldn't go in. As I fell behind my peers, I became convinced I couldn't 'do' maths. Like many others, I equated 'doing' maths with being able to memorise facts and quickly apply rote methods and, because this didn't come easily, I assumed I just wasn't a 'maths person'.
>
> These negative feelings would probably have defined my relationship with the subject but for a couple of outstanding maths teachers I encountered. These were teachers who encouraged me; teachers who taught me to find creative ways to solve problems. They showed me that there was much more to mathematics than learning (or not

learning) times tables and calculation algorithms, and, bit by bit, they developed in me a real love of the subject.

I went on to do Maths and Further Maths at A-level and then to study maths at St John's College, Oxford, before training as a teacher.

Unfortunately, in our test-driven culture, far too many young people never recover from this early experience of maths. Negative perceptions are reinforced at every stage. Whilst recent shifts in maths teaching are a welcome step, primary maths teaching is still driven by the assessment system.

What's wrong with the current model of assessment?

There are numerous problems with the current model of primary maths assessment, which have been well-documented elsewhere. We want to highlight four particular issues which we believe can be addressed through the use of observation.

i) **The obsession with product and undervaluing of process**. Whilst mathematical reasoning is directly linked with long-term success in mathematics (see for example Jo Boaler's research in *The Elephant in the Classroom*), our test-based model prioritises getting the right answer. In the 2016 KS2 tests, for example, only one mark out of 110 was awarded for a written explanation.

This approach can mask significant misconceptions. Getting the correct answer 24 x 10 = 240 might conceal a shallow understanding, leaving the student unable to solve 240 x 10 or 2.4 x 10.

ii) **The reduction of complexity**. Standardised testing reduces complexity at every level, whilst the very nature of mathematics revolves around the ability to find patterns and generalise without losing complexity.

We also lose complexity when evaluating achievement. A range of mathematical domains, usually dominated by arithmetic calculation, is reduced to a single number. By the time this is assembled at school level, the data is almost completely meaningless, having been stripped of any useful link to children's understanding.

iii) **The promotion of memory over cognition.** A clear example is the proposal to introduce a 'times tables' test during KS2. This is not to say that multiplication facts are unimportant, but the moment you introduce a high-stakes test of something, you send a signal that this is what matters most. It will result in more pressure to rote-learn these facts, rather than to understand the principles or how to use multiplication in practice. Worst of all, we send a clear signal to some pupils that they will never succeed at mathematics.

iv) **Process reduced to sterile procedure**. Having been almost entirely stripped out, process re-enters in its most sterile form – as prescribed method. That marks can only be given for working if DfE-approved methods are used sends a clear signal to young mathematicians: "Follow approved methods and don't think for yourself."

How does assessment through observation help?

Observation allows us a powerful tool to assess the sophistication of children's approaches to problem-solving and identify opportunities to develop this further. An excellent example is given in Susan Lamon's book *Teaching Fractions and Ratios for Understanding*. Each chapter begins with examples of children's work, which are then analysed in terms of their understanding of fundamental mathematical concepts and their approach to problem-solving.

Observation allows us to capture the complexity involved in problem-solving over time. This means using descriptive, rather than numerical or tick-box, records and having a clear system for identifying and recording information about children's approaches. This is useful for formative assessment but can also play a summative role, for example as a basis for providing descriptive feedback for parents about a child's progress, or to school management about overall progress across a class.

Observation allows the teacher to focus on what is meaningful and to communicate this to students. It can show students clearly that their reasoning matters.

Finally, it removes the need to use 'approved' methods because the sophistication and suitability of the child's approach to the problem are fundamental. This allows a variety of methods and approaches to be compared and contrasted.

In short, observation deepens our understanding of children's mathematics.

Observation in the primary classroom – is it practical?

In this section, we want to look briefly at three approaches, drawn from experience in our own Y5 class over the past year and developed in co-operation with teaching colleagues. The first is the most pure approach to observation but also the most time-consuming.

Option 1 – Individual Observation. A lot of information can be gained through direct observation of a child working independently on a short problem-solving task. It is important to select a task that requires students to explore their conceptual understanding of a problem and demonstrate key skills and knowledge. Jo Boaler's guidelines on developing '*rich mathematical*

tasks' (chapter 5 of *Mathematical Mindsets*) proved invaluable, as did examples from Susan Lamon's book and the NRich website.

We debated whether the teacher should remain silent but finally decided that a certain amount of questioning is needed to obtain the maximum amount of useful information from the process, though the questioning needs to avoid 'leading' the student's thinking. Our questions tended to be along the lines of seeking clarification or explanation, or prompting the use of practical or pictorial approaches ("Can you draw a picture to show me that?").

Individual observation provided us with significant insight into children's thinking but was incredibly time-consuming. Alternative tasks had to be provided for the rest of the class while particular children were being observed. We used this sparingly to gain a deeper understanding of where particular children were in their thinking.

Option 2 – Whole-Class Observation. For this, we designed a series of longer problem-solving tasks, connected by a theme, and used these as the main class work over the course of a week. One example of this was a week-long 'Pirate Treasure' investigation during Year 5. Our term's work had focused on factors and multiples, using these to understand prime and square numbers, mental methods for multiplication and division, equivalent fractions, and addition and subtraction of fractions. At the start of the final week, the children came across a coded note in an envelope marked with a skull and cross bones. This was the first in a sequence of clues, each leading to the next, and which all had to be solved to find the treasure. The children worked in pairs to try to crack the code. Each code was designed so that, once the code itself had been broken (and key mathematical understanding demonstrated) there was still a substantial task in decoding the message. This allowed

teaching staff to circulate, questioning individual children as they worked in pairs.

By the end of the week, we could confidently say we had our most accurate assessment data, having interrogated the children's understanding personally. The children enjoyed it so much that they didn't realise they were being assessed, and many asked us when they would have to do an assessment week like the other classes!

Option 3 – Journaling. This turned out to be the most efficient, and though not strictly speaking observation, it gave us similar insight. We encouraged children to express their mathematical reasoning through regularly annotating their work. To begin with, this was often forced and in response to direct questions or sentence starters given by the teacher. As the children became more familiar, they began to spontaneously contribute their thinking in written form as part of answering questions and solving problems.

This journaling uncovered key misconceptions that would have gone undiscovered based on their answers and working alone. It also contributed significantly to the children's ability to explain their reasoning verbally and in writing.

Conclusions

Observation has resource implications in terms of adult time but it provides a much more secure basis for assessing children. There is still a role for written tests, whether through a test bank which teachers can use as they see the need or for occasional sample monitoring of teacher assessment, but it should not be the main assessment mode.

Section 5: Portfolios

One important but neglected way to evaluate and evidence learning is through a portfolio. Portfolios can take various forms: a physical collection of written work and drawings, sound recordings, collections saved electronically. These collected samples of work help continuity from one teacher to another, and provide a means of reflecting on learning for pupils, staff members and parents. They can be used formatively and summatively. Students should be involved in selecting contents, a process which develops their understanding of quality.

We start this section with Tony Eaude's study of portfolios for assessing the humanities, broadly understood to include the arts and philosophy as well as history, geography and religious education.

The second set of extracts derives from Kathe Jervis' research visit to a Massachusetts school with well developed practices of engaging students.

In the third brief extract, Grant Wiggins insists on clarity of purpose, since the aim will affect what is selected for inclusion.

Finally, we present some notes of guidance which were written for students. These were intended to reduce teacher workload but also to enhance students' sense of purpose and understanding of quality criteria.

Assessing the humanities in the primary school using portfolios

Tony Eaude

While 'the humanities' are often equated with History, Geography and Religious Education as discrete subjects in England, other countries, including Northern Ireland, Scotland and Wales, organise the primary curriculum through broader areas of learning. I suggest that the humanities should be seen as whatever enables children to understand themselves, other people and the culture(s) in which people live. As such, the humanities are closely linked with spiritual, moral, social and cultural development; they are concerned with what it means to be human and with enabling children to become active, thoughtful and compassionate citizens.

This suggests a broad interpretation of the humanities, crossing disciplinary/subject boundaries and involving much more than factual information. These areas of learning can be explored and expressed through the creative arts, literature, drama and philosophy, as well as history, geography and RE.

In the current policy climate, such areas of learning have been increasingly marginalised. One may, reasonably, think that it is best not to assess such areas of learning if this entails regular testing. However, assessment should not be equated with testing. The current emphasis on testing is based on the questionable assumption that all types of learning can be measured or graded. Tests usually assess propositional knowledge and outcomes, whereas much of the knowledge associated with the humanities is procedural and personal/interpersonal, including ways of working, attitudes, values and dispositions.

Assessing areas of learning such as the humanities and the arts in terms of an accumulation of knowledge has limited validity. There is a danger that testing, especially if it is 'high-stakes', will constrict children's imagination and creativity, lead to de-motivation and add to children's stress and teacher workload. However, such challenges do not mean that one should *not* assess children's learning in the humanities; this would devalue their importance and further marginalise them. Instead, more appropriate and humane methods of assessment are needed, based on key principles of formative assessment.

Key principles in using assessment to enhance young children' learning and motivation

The first principle is that assessment must be aligned with aims. One should assess what one values. Therefore, if one values the 'whole child', one should assess, but *not test*, the whole child. This implies finding suitable ways to assess skills, understanding and qualities, rather than only propositional knowledge, and of assessing cross-curricular learning meaningfully.

A second principle is that assessment should be primarily formative and help guide and enhance future learning, rather than just assessing what children have learned. This does not exclude the use of the same information to communicate to parents what a child has achieved.

Where possible, the assessment of young children should include divergent elements, to establish and build on what they know and can achieve. It should not be restricted to what the curriculum predetermines or what we plan will be learnt. One should try to assess process as well as outcome and different types of knowledge, rather than just propositional knowledge.

It is best to integrate assessment with learning and teaching. It should also be contextualised if children are to understand the task and what is required,

whereas tests are normally decontextualised, and sometimes deliberately so. For instance, to assess skills and qualities such as teamwork and persistence, adults need to provide children with opportunities to demonstrate these in practice.

To be valid and reliable - a significant challenge with young children - assessment should be based on several perspectives. This implies using a range of different means of assessment, over time; and drawing on judgements and insights from a range of adults who know the child in different settings, including parents/carers and support staff.

To help enhance future learning and motivation, assessment should highlight children's past achievements, involve the children, and identify next steps. It should, as far as possible, be child-led to create a sense of children's voice and agency; and encourage them to be proud of what they have achieved. In other words, children should have a say in what is assessed, and how, and believe that assessment is worthwhile in that it helps to enhance their learning. Self- or peer-assessment can help children to understand what good work looks like, identify strengths and shortcomings, and know how to improve.

There is a strong argument for using spoken or written comments rather than scores or grades. Moreover, assessment should not involve grading or measurement of what cannot sensibly be graded or measured, especially where this might have perverse consequences. For example, trying to score an individual's ability to work in a team may easily skew how individuals act in the attempt to gain a high mark.

Assessment should not cause distress to the child. Methods must be manageable for both children and adults, and not take up too much time, without becoming so simple that the results are meaningless. What is

manageable will depend to some extent on the children's age and their ability to organise themselves, but also on expectations and guidance from adults.

The practicalities, benefits and challenges of adopting a portfolio-based approach

By a portfolio, I mean a collection of work of different types, gathered over time and so cumulative and contextualised. It will usually involve work in progress and finished work of different types. This may include plans, mindmaps, notes, finished work and reflections; it can encompass written work and drawings, photographs, audio recordings and video (eg artefacts, presentations, performances).

Ideally, a portfolio should:

- record work or celebrate achievements in and out of school, with descriptions of activities and certificates
- show what the child can do both individually and in a group
- be largely child-led, but including comments from adults (teachers, support staff, parents/carers and others).

It should not be restricted to what the curriculum demands. Portfolios have the potential to provide evidence of achievements over time and to enhance young children's learning and motivation. They help children to reflect on previous successes and possible next steps and become a source of pride for children and other people.

Portfolios may be related to a specific area of learning, module or period of time, though my general preference is to include a wider range of items. Portfolios can help in recording elements of process and output and assessing a range of skills and qualities. A portfolio of work associated with the

humanities can help in assessing not only content and progress in subjects such as History, Geography and Religious Education. It can also provide evidence of children's ability in writing, drawing and talking, as well as problem-solving and co-operation. Moreover, such an approach helps to value children's existing 'funds of knowledge', including those not traditionally valued by schools.

Organising a portfolio

There is no single way of collecting and organising a portfolio. Children will need guidance on what to include. Teachers may expect children to include specific pieces of work to evidence specific subjects or topics. However, children should be encouraged to add what they believe to be appropriate, within limits, if they are to retain ownership of the portfolio. When well established, a portfolio-based approach helps reduce marking.

One difficult challenge is how to ensure that work does not get lost or damaged. Compiling a physical collection can restrict size and format, and bits of paper can get lost. ICT makes it easier to store a wider range of written work, drawings, scanned documents, photographs and recordings than traditional portfolios.

Children should be expected to add to the portfolio weekly or fortnightly - and out of school as appropriate, with adults also adding items. It can then be reviewed and pruned on later occasions.

A portfolio-based approach is easier for older primary-aged children (from about 7 years old upwards) to manage independently, but most young children should be capable of organising material with support at first.

Conclusion

This article has challenged the conflation of assessment with testing and argued against trying to measure what cannot appropriately be measured. Placing children's agency and voice at the heart of assessment helps to engage and motivate children by promoting pride in their achievements across a wide range of areas of learning, both in and out of school. Such an approach has the potential to give children more responsibility for, and control of, assessment and to reduce teacher workload.

A portfolio-based approach helps integrate assessment and learning, showing how skills and understanding have been applied in particular contexts. It provides evidence of process and outcome over time, encourages reflection on previous work and consideration of next steps. It particularly lends itself to assessing the humanities, in a broad sense, and the arts and children's talk - aspects for which tests are likely not to be valid, reliable or meaningful.

Assessment by portfolio

Kathe Jervis

Describing her research visits to a school which had pioneered the use of portfolios, Kathe Jervis identifies *assessment as inquiry* as its guiding principle, not assessment as measurement. As one teacher put it:

> The development of portfolios came out of all the work we did with curriculum over 15 years - we knew we couldn't tap the deep skills kids had with any kind of testing, standardized or otherwise. We knew we weren't getting all we could. My impetus was to validate kids' voices.

A staff committee had summarised key principles in a handbook, drawing on the writings of Elliot Eisner:

> Education out to promote the exploration of ideas, leads to a better, more satisfying life outside of school, moves away from extraneous reward systems, and engages children in formulating their own problems to solve. Further, educators should engender a tolerance for ambiguity and provide opportunities for children to express their imaginings.

Part of the aim was to validate the different forms of expression and activity neglected in written tests.

Dimensions of learning

The staff developed the idea of four *dimensions of learning* to shape their practice, and to underpin discussions with children and parents. These dimensions also ensured broad and rounded criteria for evaluating whether children are reaching a good standard.

i) *Acquisition and Application of Knowledge* covers concepts, information, processes, and skills.

ii) *Communication* includes a standard for written work, oral reports, projects (construction, slideshows etc) and performances (drama, music, etc.) where the emphasis is on effective presentation of information, ideas, and/or feelings.

iii) *Attitudes and Approaches* attempts to describe the process of learning to learn, the stance children take toward their work.

iv) *Reflections* considers making children conscious of their own learning processes.

Attitudes and behaviours raises expectations such as:

- engagement in activity
- persistence and risk-taking
- taking a deliberate and thoughtful approach.

A key aspect is 'using thinking abilities to solve problems, make decisions, and examine issues' examplified as follows:

- shows interest in problem posing and / or issue examination
- uses various problem-solving and /or decision-making strategies
- collects data from a variety of sources
- asks questions
- seeks evidence
- seeks alternatives and evaluates them critically.

The active involvement of children

One teacher reports how she deliberately sets out to engage children in assessment. At the start of the year, she asks her class to brainstorm the question: "How can you prove you have learned something?" Children suggest a range of modes such as: write it down, tell someone, make something, use video, computers, photographs.

Children collect samples of their learning during each term. They are required to make a recording of their reading at least once a year, which are used to demonstrate progress over time and to pass on to the next teacher. On Fridays, the children fill out reflection sheets to record in their own words what they have done. The four dimensions are considered, and at the end of term each child must include something for each dimension.

The skeleton

Kathe Jervis recalls how, on one day, a boy succeeds in putting together a full-size human skeleton, borrowed from the school nurse. This boy, despite his limited literacy skills, is using several reference books open at skeletal diagrams.

> Rising to the full height of his 7-year-old stature, hands on his hips with a bit of bravado, almost posing beside his accomplishment, he tells me, "I wanted to be the first to put the skeleton together and I was."

The teacher rushes across with a camera to document this complex piece of second-grade science work:

> "During the week a child or I can request a photo of some learning that seems significant, or on Friday a child can ask me to photograph

> work which seems important in retrospect. Later, with photos in hand, students describe in writing what is important about that work, and together we figure out what we can learn about them from the photo and the description."

She adds that before the photo goes into his term's portfolio, she will interview him to find out more and to clarify the guidance she can give.

The end of the year: more than a folder

At the end of the year, the children review their three termly portfolios and prepare their "pass-along portfolios" to take with them to their next teacher. As one teacher comments:

> Although I require some types of work, the child has the final say. We might negotiate on a particular piece. As the children look through their collections, discussion centres around the balance between enough and too much. I ask each child to answer the question, "Who are you? What do you want someone, specifically your parents and next year's teacher, to know about you?" Much self-discovery takes place and each child's portrait begins to emerge for the pass-along portfolio. Each child writes a letter of introduction to accompany the portfolio.

The following items must be included:

- a letter to the next teacher
- self-portraits from the first week of school (art)
- some self-chosen photographs with children's explanations attached
- photos such as "What I want people to know about me"
- drafts of an illustrated story

- self-chosen artifacts from field trips (social studies and science)
- a reflection on their Reading Project.

The teacher explains that last year, the children's task for the June Reading Project was to figure out some way of transforming their favourite book of the year into a performance, an art project, or a book report.

Growth over time: mysterious and elusive

As Kathe Jervis explains, looking through portfolios as they develop shows some exciting breakthroughs in children's work:

> moments when they begin to have more insight, when their writing all of a sudden takes account of conventions, and when their passions show through. Miraculously, without explanation, these breakthroughs occur at different times for different children, reminding me that much about human learning and how it happens cannot be captured at any one point in time.

Although it is tempting to use these to compare different children, she feels that this cannot do any individual child justice:

> Yet standards of knowledge that schools and society attempt to impose invite these comparisons as well-intentioned educators strive for excellence, accountability, and a well-prepared world-class work force.
>
> In contrast to a portfolio, a single numerical test score measures the *trivial*, robs learning of its *rich complexity*, and reduces the possibilities for children to demonstrate their learning.

Portfolios as evidence

Grant Wiggins

The late Grant Wiggins supported portfolios of students' work because they are able to bring together diverse evidence of learning. They can involve students more effectively in self-assessment, and form a basis discussions with students and parents.

He also warned about the problems if portfolios are not well planned:

But just what is a portfolio? Uncertainty abounds about just what should go in a student portfolio, who really should own it (the student or the teacher), and what value, if any, there is in passing portfolios along to the next teacher or school. This confusion stems in part from a failure to think through the purposes of the portfolio.

In fact, there are very different possible purposes, including:

- as a showcase for the student's best work
- as a showcase for the student's interests
- as a showcase for the student's growth
- as evidence of self-assessment and self-adjustment
- as evidence enabling professional assessment of student performance, based on a technically sound sample of work.

A portfolio designed to show evidence that a particular performance standard has been reached is very different from a portfolio designed to show progress over time - from novice to expert - and is very different from a portfolio designed to show the student's interests and abilities as judged by the student. In each case, we would expect to see very different samples of work.

Teachers often worry about the workload but this is because there is not a sharp enough sense of purpose.

In all cases, far fewer samples are needed than people typically imagine. A huge collection of student work is not necessary to show that a particular standard has been reached or that a particular interest exists and has been developed. A few papers, projects and tests are likely to be sufficient to certify that a student is now advanced or is talented. A portfolio is fundamentally a sample of work, regardless of its purpose. It is not a file cabinet or exhaustive collection of artifacts.

Thus the distressed cry of "Where and how do we keep this massive amount of stuff?!" is indicative of some muddled thinking about purpose. Once two teachers have certified that Susie can write at the intermediate level, based on six samples of writing in the fourth grade, they no longer need more than one or two representative pieces of work as backup to the judgement in the event of a future dispute.

Even a portfolio designed to chart progress over time with a strong emphasis on self-assessment need not include all or most work. A careful pruning based on specific rubrics will likely result in a far smaller but telling collection of artifacts than many teachers commonly deal with.

Guidance notes for students: Persuasive Essay

Attach one piece of writing that demonstrates your ability to write persuasively. Include assignment sheets if you have them.

People who read this essay will look for evidence that you can:

- engage the reader by establishing a context, creating a point of view or persona, and using other appropriate techniques to develop reader interest
- include a controlling idea that organizes your writing and makes a clear and logical judgment
- organize your writing in a way that is appropriate to the needs and interests of a specified audience
- arrange details, reasons, examples and/or anecdotes effectively and persuasively
- include appropriate information and arguments; exclude information and arguments that are irrelevant
- anticipate and address reader concerns and counter-arguments
- support arguments with detailed evidence, citing sources of information.
-

Describe the assignment that prompted this work (attach other pages if needed).

What makes this work a good piece of evidence for this entry?

Guidance notes for students: Communicating Information

Attach evidence of your ability to make an oral presentation to share information you've gathered. Evidence may include planning notes, an outline, speech notes, or audio or video tapes. In addition, you must include teacher, peer, and self evaluations that demonstrate your ability to:

- ask appropriate questions
- respond to the questions of others
- paraphrase and summarize to increase understanding
- listening responsively to others' points of view
- use language that is simple and appropriate for communicating
- speak audibly.

Describe the assignment that prompted this work (attach other pages if necessary).

Section 6: Authentic and holistic assessment

There is a frequent tension in assessment between validity and reliability. Validity answers the question 'are we really assessing what we intend?' whereas reliability is about consistency of results.

Because a national standardised test such as SATs needs to be fair to all, there is a tendency to opt for rigid marking schemes, fragmentary questions, and unambiguous answers, even if this means a loss of validity. It is rather like assessing the ability to swim 20 metres through a multiple choice test - consistent perhaps, but scarcely valid.

The late Grant Wiggins worked for many years on the development of *authentic* assessment which comes close to the complexity and messiness of real life situations. He regarded this not as a total replacement for tests of knowledge, but as complementing them. His argument was that understanding and judgement are needed as well as knowledge when evaluating creativity. We begin this section with extracts from his general argument.

This is followed by his rubric for assessing creativity. Rubrics provide important guidance to teachers and learners, though they should not be treated dogmatically and always require professional development to form a consensus on what such a performance will look like.

Eddie Playfair, principal of Newham Sixth Form College, argues for a reform of assessment in the 16-19 stage. After discussing the principles of the International Baccalaaureate (IB), he focuses on the Extended Project Qualification (EPQ) which can be used by any school or individual to complement their specialist studies.

We continue with David Leat's summary of the the EPQ, which explains how it works as a complement to A-level studies, and why it is valued by universities as a sign of wider learning and readiness for advanced study.

Finally, we report on how assessment in the form of 'rich tasks' was developed within the New Basics curriculum in Queensland, Australia. The rationale, and some interesting examples, are presented which will hopefully provide a stimulus for teachers in all stages and subjects.

Swimming lesson on dry land

Assessing creative learning

Grant Wiggins

Much of the work of the late Grant Wiggins consisted of seeking more creative alternatives to high-stakes testing in the USA. He argued that creativity can be found in solving mathematical problems. Quoting Polya (1957), he saw the essence of mathematics as "the ability to solve problems - not merely routine problems but problems requiring some degree of independence, judgement, originality, creativity."

For Wiggins, creativity was not just novelty but also required a sense of purpose or appropriateness:

> Various researchers on creativity have defined it as "the ability to produce work that is novel and appropriate." That gets it just right, I think: being merely imaginative, offbeat or inquisitive may be delightful but as educators we should not regard it as sufficient... There has to be an "appropriate" impact – whether in joke telling, fine art, philosophy, engineering or athletics...
>
> The point is generalizable. Socratic Seminar, working with primary source texts / artifacts in history, playing football well as a team or developing a jazz guitar solo elicits creative learning because thoughtfulness is demanded by the task.

Creative activities place very different demands on the teacher:

> As video games so clearly illustrate, creative learning demands very little "teaching" as long as there are clear challenges, good feedback, and choices for the learner to make.

It involves a sense of play, which brings about the highest quality learning:

> We are only truly learning when we try to apply (and make sense of that using) what was taught. In this way, academic work is really no different from guitar or hockey: We haven't begun learning unless we *play*.

Consequently, assessment needs to look not at originality in the abstract but at the achievement of particular aims in specific situations:

> Did the performance work? Was the purpose achieved – even if in an unorthodox or unexpected manner? Creativity can only be evoked and developed if we assess for such impact... The point of performance assessment is not to have students merely emulate the form and content of past performances and performers, but to emulate the best effects, eg the ability to persuade an audience, satisfy a client request, or solve a problem...

Questions of impact go beyond simply pleasing the teacher:

> Unless we highlight *impact* criteria, the student in fact has no genuine performance goal other than to please the teacher or mimic orthodox approaches. "Is this what you want, Mr Smith?" is a vital sign of the failure to teach students that performance criteria are not about custom or teacher preferences but about what actually tends to be *novel and appropriate* – i.e. what really works.

This means that students must be introduced to "many diverse models of excellence and non-excellence at meeting performance goals creatively."

Feedback too must relate to particular aims and situations, and look at the adequacy of the performance:

> Truly helpful feedback attends to the ultimate desired outcome, and gives you information about how you did against that bottom-line goal. We hit the tennis ball and see where it lands, we give a speech and hear audience reaction as we speak, we design an experiment and check the results for error... Feedback is merely the answer to the question: What happened?
>
> Consider: People laughed at the first joke but not at the second and third joke. Why? What can I learn from the feedback about how to make them laugh at all three?

Consequently, feedback is different than praise or blame:

> Good job!" and "Try harder!" are not feedback. Praise and advice can certainly be useful; but valid descriptive feedback is always useful, empowering – the source of all creative learning. How would the public speaker become skilled and poised if there were never a real audience and experts merely wrote back with letter grades a few weeks later?

A generic rubric for assessing creativity (extracts)

Grant Wiggins

6) The work is unusually creative. The ideas/materials/methods used are novel, striking, and highly effective. Important ideas/feelings are illuminated or highlighted in sophisticated ways. The creation shows great imagination, insight, style, and daring. The work has an elegant power that derives from clarity about aims and control over intended effects. The creator takes risks in form, style, and/or content.

- The problem has been imaginatively re-framed to enable a compelling and powerful solution
- Methods/approaches/techniques are used to great effect, without overkill
- 'Less is more' here: there is an elegant simplicity of emphasis and coherence
- Rules or conventions may have been broken to create a powerful new statement.
- Common materials/ideas have been combined in revealing and clever ways
- The audience is highly responsive to (perhaps disturbed by) the work
- The work is vivid through careful attention to telling details and deft engaging touches
- There is an exquisite blend of the explicit and implicit

4) The work is creative. The ideas/materials/methods used are effective. A voice and style are present.

- Novel approaches/moves/directions/ideas/perspectives were used to good effect

- There are imaginative and personal touches scattered throughout the work
- The work keeps the audience mostly engaged
- There is a discernible and interesting effect/focus/message/style, with lapses in execution
- The work takes some risks in methods/style/content

2) The work is not very creative. The approach is trite and the ideas clichéd, leading to a flat and predictable performance. There is little sense of the creator's touch, voice or style here.

- The work offers little in the way of new approaches/methods/ideas.
- There is little sign of personal voice, touch or style.
- The work suggests that the creator confuses 'creative' and 'risk-taking' with 'shocking in a juvenile way'.
- There is excessive and incoherent use of different materials, techniques, ideas
- The creator may have confused great care and precision with creativity – the work is more polished than imaginative or revealing.

Assessing the sixth form 'masterpiece'

Eddie Playfair

Sixth form education in England is defined by its qualifications rather than its educational purpose. Students enrol on a programme made up of specific qualifications and their success is measured in terms of their achievement on those qualifications. The value of these qualifications is mainly an exchange value; providing access to the next stage of education or employment rather than having an intrinsic educational value.

English sixth formers' experience of assessment is therefore almost entirely driven by the requirements of nationally accredited standardised public exams organised at subject qualification level. The emphasis is on sorting and grading accurately and consistently within qualifications. This restricts the system's ability to take a holistic view of a students' learning and achievements.

At A-level, assessment is dominated by terminal, timed, individual written exams which tend to value knowledge recall in discrete chunks. Although some credit is given for 'synoptic' understanding at a whole subject level, this is still limited to within-subject knowledge and skills. Vocational diploma courses provide more scope for portfolio or assignment-based assessment and there is at least the possibility of designing substantial integrated assignments which can value a more complex combination of knowledge and skill embedded in authentic challenges in real-world settings.

There is a strong case for a broader and more coherent 16-19 curriculum capable of offering all students an introduction to a fuller range of human culture and experience while also allowing for some specialisation and the pursuit of personal interests. Such a curriculum should make greater demands

on young people by aiming for a broader level of general culture and the ability to make connections between different aspects of their learning so far. Very few currently available programmes in England do this. The International Baccalaureate (IB) Diploma is a notable exception, but it available only to a tiny fraction of the cohort.

The design principles of the IB curriculum require it to be: *broad and balanced* - with content that spans subjects, *conceptual* – focused on powerful organising ideas that have relevance within and across subject areas, and *connected* – helping students draw connections and understand the interrelationships of knowledge and experience across many fields. The three core elements; *theory of knowledge,* the *extended essay* and *creativity, activity and service* aim to challenge IB Diploma students to broaden and deepen their experience and apply their knowledge and skills. The IB system uses a range of assessment strategies to support good practice by encouraging authentic 'performances of understanding' which require creative and critical thinking. Final assessments include coursework and external exams while being rigorous and internationally benchmarked.

Sadly, this sense of purpose and coherence is often lacking in the thin programmes currently on offer to most 16-19-year-olds in England. But we don't have to ditch everything and start from scratch. In the absence of a nationally defined unifying purpose for sixth form education or the kind of curriculum and assessment which might support this, there are still things which we can do to tweak our current system to add greater meaning to the sixth form offer and to assess young people's skills and capabilities in a more rounded way.

A key tool for this could be the existing suite of research awards, not so very different from the IB's extended essay: the Extended Project Qualification (EPQ) at advanced level and equivalent to half an A-level, the Higher Project at intermediate (GCSE) level and the Foundation Project at level 1. Using these to support a common core of broadening studies with the majority of sixth formers across the board, rather than a small minority, could be a practical way to bring about a real step change in post-16 education.

A good EPQ programme allows a young person to investigate a question which interests them critically, analytically and in some depth. Their topic might be a deeper exploration of a theme being studied in one of their subjects, it may arise from the interaction of different subjects or in the spaces between them, or it may be something entirely personal and unrelated. At its best, it can be an original contribution which involves some primary research and offers a genuinely new insight. The EPQ is an opportunity for students to produce their version of an apprentice's 'masterpiece' which demonstrates their commitment and their promise and makes a tangible contribution to their community. It should be something they can proudly present to a wide audience as the culmination of their skill to date and which can provoke further reflection and discussion.

At its best, the product of such a student research project provides evidence of mastery and skill which can hold its own in the wider world and this could form part of everyone's sixth form graduation or matriculation. For today's visual or performing arts students, this evidence could be similar to their current portfolios, artefacts or student-devised productions. For students of other disciplines, it might be a student-led community project, social enterprise, publication or the more traditional written essay.

How does one assess such a 'masterpiece' or capstone project? Candidates need to present their work to an audience and be prepared to 'defend' their findings in a way a PhD candidate does in a viva. Criteria include clear communication, methodological rigour, skill, accuracy, originality, imagination and usefulness to the wider community. Sixth form teachers, university academics, professionals, employers, local residents and other students can all play a part in supporting, assessing and celebrating this work. Being on a panel, listening and responding to others and giving critical and constructive feedback can all be part of what is assessed in the round.

The EPQ is valued by universities as a measure of students' academic curiosity as well as their research and communication skills and they are the obvious candidates to help kick start a renaissance in this approach to student learning. Universities could extend and deepen their support for developing a research culture. Regional partnerships could provide training and resources for sixth form staff and students across a wide area.

Digital platforms offer a great opportunity to share and discuss project outcomes widely in the public domain. Each project can be a significant milestone for a student's development as an active, thinking contributor to human understanding and to the wider culture and continue to be commented on by others. It could be followed by further capstone projects at different points in life as an undergraduate, a postgraduate, a worker or a citizen. Both the product itself and the process of production represent educational and social 'good work' as validated by experts as well as other interested people. The sum of all interconnected and related projects could be a substantial resource for teachers and learners.

Such an approach would move us one small step away from a system which mainly values the proxies for the useful knowledge and skill which help us do good work and one small step towards valuing the good work itself as part of the individual's ongoing social contribution to making a difference in the world.

If all we want to do is to rank and grade young people's work objectively with little reference to their continuing learning journey, then there's no case for change. But we can surely be more ambitious than that for our students by aiming to connect their learning more explicitly to the world they live in and to describe how it might be useful to them and to others in future. Rather than simply being about summative judgements about an individual exam performance, the assessment process could be part of an ongoing conversation about the use of knowledge and skill in society, supporting our students' growing understanding of, and contribution to, the world.

The Extended Project Qualification (EPQ)

David Leat

The Extended Project Qualification was first introduced in 2009, following a review of post-16 qualifications in 2006. It can either be taken as a standalone qualification or part of a Baccalaureate. It is equivalent to 50 per cent of an A level.

There is a small taught component, including research skills, analysis (geared to the specific project), project management and report writing. The student has free choice of topic, but they are expected to show that it is academically relevant. They are supported by a supervisor who works under a centre (usually, school or college) coordinator.

The EPQ requires students to choose, plan, research and conduct a project under their own initiative. Learners develop an initial idea for a project, which they discuss with their supervisor. They then conduct some initial research as a basis for developing their ideas sufficiently to make a formal proposal that includes their aims, initial plans and likely format for presentation. They then complete a Project Proposal Form. The Centre Coordinator comments, and can either approve, require changes or request a resubmission.

Self-efficacy and creativity are encouraged by allowing students to choose the format of their outcome. This can take the form of a

- research-based written report
- production (eg charity event, fashion show, sports event etc)
- artefact (eg piece of art, a computer game or a realised design).

In the latter two cases, a written report must also be included.

In addition, assessment includes standardised project record forms, particularly the Production Log and Assessment Record which trace the evolution of the project idea into realisation.

On completion, the learners must give a presentation, which should be for a non-specialist audience, using media appropriate to the type of project. The presentation must include a live question-and-answer session overseen by the supervisor.

The assessment objectives are weighted as below:

Manage: Identify, design, plan, and carry out a project, applying a range of skills, strategies and methods to achieve objectives (20%).

Use resources: Research, critically select, organise and use information, and select and use a range of resources. Analyse data, apply relevantly, and demonstrate understanding of any links, connections and complexities of the topic (20%).

Develop and realise: Select and use a range of skills, including, where appropriate, new technologies and problem solving, to take decisions critically and achieve planned outcomes (40%).

Review: Evaluate all aspects of the extended project, including outcomes in relation to stated objectives and own learning and performance. Select and use a range of communication skills and media to present evidenced project outcomes and conclusions in an appropriate format (20%).

The numbers taking the EPQ have risen steadily, from 15,958 in 2010 to over 35,000 in 2016. Although some students worry that it could take time away from A-level studies, some of the most prestigious universities have described it as great asset in securing offers and during interviews.

One headteacher commented: "Taught properly, we have found the extended project is genuinely ground-breaking and transformational for both pupils and teachers.' Some of the extended projects being completed by pupils at the school include:

- Is prenatal genetic screening and testing ethically justifiable?
- Does science leave less room for free will?
- If wind turbines are economically viable, why do they currently provide so little of our electricity?
- Is climate change no longer a scientific debate but a sociological and political problem?
- What is a number?

Authentic assessment through rich tasks

Queensland New Basics curriculum

Rich tasks have to be valuable in themselves, and not distract from real learning. A *Rich Task* is a *culminating performance or demonstration or product* that is purposeful and models a life role. It presents substantive, real problems to solve and engages learners in forms of pragmatic social action that have real value in the world.

The problems require identification, analysis and resolution, and require students to *analyse, theorise and engage intellectually with the world*. In this way, tasks connect to the world outside the classroom. Rich Tasks have relevance and power in new worlds of work and everyday life. It is important that they have recognisable face value with educators, parents and community stakeholders as being significant and important.

As well as having connectedness, the tasks are also *intellectually rich*: they represent an educational outcome of demonstrable and substantial intellectual and educational value.

A Rich Task should not be seen as a short-term 'project'. It is the culmination of three years' work. Not only is the quality of the product important but also the intellectual *strategies, knowledge and skills* that are acquired by the student in the processes leading up to the completion of the task. It is crucial that tasks be rich in developmental, cognitive and intellectual depth and breadth to guide curriculum planning across a significant span of schooling.

To be truly rich, a task must be *transdisciplinary*. Transdisciplinary learnings draw upon practices and skills across disciplines while retaining the integrity of each individual discipline. This is not the same as the traditional

interdisciplinary approach that seeks links between disciplines often via thematic learning.

These processes culminate in *individual or team presentations*, based on research and problem-solving. The challenge must be meaningful to the learners; it could involve a local context, and be presented to an audience of parents or a community group.

If we really want to improve educational standards, we should use assessment which enhances learning rather than trivialising it. This could involve peer- and self-assessment which is supportive rather than judgemental, and recognition for co-operative learning in groups rather than individual competitiveness. These forms of assessment would help to foster a genuine sense of *learning community* in our schools.

In summary, a Rich Task:

- is an integrated intellectual and linguistic, social and cultural practice
- represents an educational outcome of demonstrable and substantive intellectual substance and educational value
- is transdisciplinary
- draws on a range of operational fields of knowledge
- is problem-based
- connects to the world beyond the classroom
- has face value for educators, parents and community stakeholders
- has sufficient intellectual, cognitive and developmental depth and breadth to guide curriculum planning across a significant span of schooling
- enables flexibility for schools to address the local context
- has reasonable workload expectations for teachers.

Rich Tasks for 11 year olds

Multimedia presentation of an endangered plant or animal

Students will investigate a threatened plant or animal and the extent to which it is at risk. They will use this investigation to take constructive action and create a persuasive and informative multimedia presentation.

Oral histories and diverse and changing lifestyles

Students will explore change in, and diversity of, modern lifestyles, with particular reference to the nature of work, by recording oral histories from various members of their own community, including people in a variety of cultural groups. They will use the oral histories as the basis for a media presentation that portrays significant changes in work practices in the past and predicts how work practices might change in the foreseeable future.

A celebratory, festive or artistic event or performance

Students will work within teams, in different capacities, in planning, organising, creating and performing in a celebratory, festive or artistic event or performance that takes place at or outside the school.

Travel itineraries

Students will design alternative itineraries of interest to a party comprising the student and an exchange student, and to be accompanied by an adult. They will identify a range of issues including transport options, tourist attractions and sites of historical and cultural significance. They will present costings and reasons for their choices.

Rich Tasks for 16 year olds

Improving health and wellbeing in the community

Students will work with a local community to develop a plan for improving an aspect of the wellbeing of this community and then enact the plan, modifying it as necessary. They will evaluate the level of success they experience in enacting their plan and, where necessary, recommend future actions.

National identity: influences and perspectives

This project involves the planning, production and presentation of a powerful, filmed documentary including information gleaned from research and interviews with people from different cultural backgrounds. Students will demonstrate knowledge and understanding of different influences and perspectives on national identity.

Opinion-making oracy

Students will make forceful speeches on an issue of international or national significance to different audiences.

The shape we're in

Students use mathematical skills to investigate alternative shapes and/or dimensions for at least one container, a domestic object, a mechanical device and an object from nature. They then present an alternative design for one of these, explaining the mathematics.

Pi in the sky

Students will demonstrate an understanding of different mathematical approaches used to frame and answer questions about astronomy asked by

cultures from three different historical ages. For each culture, they will immerse themselves in one such question as well as the ways in which the culture used or developed mathematics to frame and answer the question. They will then present one of three lessons, chosen at random, to communicate the essential ideas and techniques of the mathematics of the situation.

International trade

Based on knowledge of the way in which international trade occurs and is reported as well as knowledge of the needs and wants of another culture, students will identify and provide a detailed analysis of an export opportunity. They will take advantage of their skills in a language other than English to present a talk and supporting literature to promote this export opportunity to different buyers and backers.

Final comment : The grass roots speak

John Coe (National Association for Primary Education)

Three groups of people are directly affected by the assessment of primary education. Firstly, there are the children. All we do is for them, their present and future needs, and their wellbeing and happiness. Then there are the parents and carers who are better placed than anyone to judge the impact on their children. Lastly there are the teachers who work with the children and the parents every day.

The most affected are the children whose study of history is reduced to 30 minutes a week, the parents who deal with the crying child at bedtime because the weekly spelling test is tomorrow and the teachers who, against their professional judgement, emphasise and coach for the test results which are the measure of their competency. These are the grass roots of education and they are caught up in a politically motivated system beyond their control.

Teachers and heads have long fought against the system, researchers have written reports, conference speakers have been applauded as they produced evidence of the damaging effects of the dominance of testing. All to no avail: the testing juggernaut grinds implacably on and the political imperative is that the engine which drives education must be that of market forces as if the schools are no more than competing supermarkets.

It was in May 2016 that the parents of young children first came together to voice their deep concern and wish for the reform of the assessment system. Several thousand families kept their children away from schools in an unprecedented protest against national testing. The message to the Government was enough is enough, stop the incessant testing which is

hurting our children and find another way of assessing educational progress. This is a new and potentially powerful chapter in the story of state education.

A campaign, More Than A Score (MTAS), intent on reform, was quickly organised and this has grown in strength. Three parent-led organisations, Let Our Kids Be Kids, Rescue Our Schools and Save Our Schools, are joined by the largest teachers' union the NUT (now part of the NEU) and other voluntary associations reflecting the consensus of views among teachers, parents and other education professionals.

Testing every individual child in order to judge the effectiveness of teachers and schools is deeply flawed and has negative effects on the quality of education. It focuses the energies of pupils and teachers on achieving success in a narrow range of subjects: the school curriculum is dominated by mathematics and English, distorted by the need to make them testable. When schools are judged primarily on test results, pressure and stress builds up on pupils and teachers alike and the system becomes punitive.

National standards should be evaluated by the testing of a sample of children spread over a number of schools, as already happens with science. This would prevent any undesirable backwash into the curriculum.

The voices of those calling for change are wide-reaching and growing louder. For too long the needs of external testing have dominated pupils' entire experience of school. Successive governments have failed to acknowledge the damage caused by the over emphasis upon test results and have refused to ask the crucial question: what kinds of assessment create the conditions for young people to thrive in an uncertain and innovation-rich world? It is time for other energies stemming from the grassroots of education to answer that question.

Beyond the exam factory

Assessment in English schools is not designed to help children learn. Its main purpose is to police schools and teachers, and it does untold damage in the process. It causes stress to children, demoralises teachers, and provides little useful information to parents. It narrows the curriculum and penalises schools in the most disadvantaged areas.

This book was put together to open up real alternatives. It draws on a wealth of experience and expertise over many decades, in England and internationally. It presents examples of a wide range of assessment methods which have been eclipsed or forgotten due to the pressures of 'accountability'.

Primary schools are a particular focus, given the current crisis, but examples come from - and are relevant to - all age groups. We hope it will lead to widespread discussion among teachers and heads, parents, school governors, politicians and the general public.

Children are more than a score

Beyond the exam factory: Alternatives to high-stakes testing published by More Than A Score

Author details, acknowledgements and references

We are grateful for all the authors' permission to republish various articles in an abbreviated form. We also wish to thank the editors and publisher of the journal Forum where some more extended versions were published in August 2017.

Tim Sanders, cartoonist at Common Knowledge, generously agreed that we should re-use his drawings, as did Ros Asquith for her cartoon placed at the end of Margaret Clark's article.

Introduction

Much of the history concerning the 1988 Education Reform Act (National Curriculum and assessments, competition between schools) and the creation of Ofsted can be found on http://www.educationengland.org.uk/

Many books have been written about this period, among which:

Sally Tomlinson (2005) *Education in a Post-welfare Society*

Denis Lawton (1994) The Tory Mind on Education 1979-94

A good analysis of curriculum changes, including Michael Gove's reform, can be found in:

Terry Wrigley (2014) *The politics of curriculum in schools*. http://classonline.org.uk/pubs/item/the-politics-of-curriculum-in-schools

Part A

Section 1: Assessment and the accountability machine

Ofsted inspection and the betrayal of democracy (Michael Fielding)

Michael Fielding is currently Chair of the Editorial Board of FORUM and Emeritus Professor of Education at UCL Institute of Education.

The original article 'Ofsted, Inspection and the Betrayal of Democracy' was published in 2001 in the *Journal of Philosophy of Education*, 35(4).

A malediction upon management (Fred Inglis)

Fred Inglis is Emeritus Professor of Cultural Studies at the University of Sheffield. A long-standing contributor to the Nation, the New Statesman, and the London Independent, he is also the author of numerous books.

The original article 'A Malediction upon Management' was published in 2000 in the *Journal of Education Policy*, 15(4).

The illusions of measuring linear progress (Reclaiming Schools)

Reclaiming Schools is a network of researchers formed to provide reliable and accessible knowledge to the NUT for its campaigns. Its website is www.reclaimingschools.org

The flawed assumption of smooth linear progress was exposed by Education Datalab *Seven things you might not know about our schools.* https://educationdatalab.org.uk/2015/03/seven-things-you-might-not-know-about-our-schools/

The analysis in sub-sections 1-6 is underpinned by data on the Reclaiming Schools blog, indexed at https://reclaimingschools.org/2016/10/10/links-to-posts-on-primary-testing/

Section 2: Assessment and the developing child

Homo Sapiens 1.0: human development (Pam Jarvis)

Pam Jarvis PhD is a chartered psychologist and a historian and has qualified teacher status. She has numerous publications and is currently Reader in Childhood, Youth and Education at Leeds Trinity University.

Baseline testing: science or fantasy? (Terry Wrigley)

Terry Wrigley is editor of the international journal Improving Schools and Visiting Professor at Northumbria University. He is one of the coordinators of the Reclaiming Schools network (www.reclaimingschools.org). He has a keen interest in curriculum and social justice. His most recent book is Living on the Edge: rethinking poverty, class and schooling (co-author John Smyth).

The research published here was undertaken in conjunction with the Better Without Baseline network, and has been published in various versions including *Radical Statistics* (2016) and *Primary First* (2017).
Much of the data for this article can be found via https://reclaimingschools.org/2016/10/10/links-to-posts-on-primary-testing/

Democratic alternatives to early assessment (Guy Roberts-Holmes)
Guy Roberts-Holmes' main research interest is in the role of data and its effects upon primary and early years education pedagogy, curriculum and governance. He coined the term 'the datafication of pedagogy' within early years and primary education to critically analyse the ways in which data are increasingly used to govern, regulate and discipline education. His recent research into reception baseline assessment with Alice Bradbury ('They are Children ... Not Robots, Not Machines', commissioned by the NUT and ATL) was awarded the BERA/SAGE Research Impact Prize 2016. The arguments in this article are expanded in the authors' new book *The Datafication of Primary and Early Years Education* (Routledge2018)

Many examples and texts relating to the work of Malaguzzi in Reggio Emilia are contained in the book *Loris Malaguzzi and the Schools of Reggio Emilia* (ed. P Cagliari et al., Routledge 2016)

Section 3: General proposals

Assessment - what we stand for (More Than A Score)

More Than A Score is a broad coalition which includes teacher trade-unionists, parent groups, academic researchers and a range of professional and curriculum organisations, including various bodies concerned with early childhood. Full details of member organisations can be found at https://morethanascore.co.uk/who-we-are/

Some modest proposals (Terry Wrigley)

Terry Wrigley is editor of the international journal Improving Schools and Visiting Professor at Northumbria University. He is one of the coordinators of the Reclaiming Schools network (www.reclaimingschools.org). He has a keen interest in curriculum and social justice. His most recent book is Living on the Edge: rethinking poverty, class and schooling (co-author John Smyth).

Assessment in English 3 to 11 (John Richmond)

John Richmond has been an English teacher in London, an advisory teacher for English, an officer on two national curriculum-development projects in language education, and a local authority adviser. He has been a commissioning editor in educational television. He has published books and articles on English teaching and the role of language in learning. It draws substantially on chapter 12 of *Curriculum and Assessmment in English 3 to 11: a Better Plan*, published in 2017 by Routledge.

National tests in Denmark (Jakob Wandall)

Jakob Wandall has a background as a researcher in social sciences, evaluation and education. From 1994-2011 he was Head of Division/Chief Adviser in the Danish Ministry of Education. He was the project manager (2006-2011) on the development and implementation of the national testing system. In 2011 he co-founded NordicMetrics, and he is an external lecturer at the University of Aarhus (Department of Education). The author is currently providing consultancy advice for assessment reform in Wales. The text presented here is an abbreviated version of an article in the *Journal for Applied Testing Technology* (2011).

Part B

Section 1: Formative assessment

Science inside the black box (Paul Black and Christine Harrison)

Paul Black and Christine Harrison were part of a larger team from King's College London engaged since 1998 in a major exploration of classroom assessment practices to encourage formative assessment. They undertook a particular investigation and development in collaboration with science teachers. The booklet *Science inside the black box* is one of a set of subject-related booklets available from GL Assessment (www.gl-assessment.co.uk)

Verbal feedback (Flora Barton)

Flora Barton is headteacher at Crowmarsh Gifford CE Primary School, Oxfordshire. Here she describes one aspect of the rich assessment practices developed by staff at this school.

Denmark: learners setting goals (Kirsten Krogh-Jespersen, Anne Birgitte Methling and Andreas Striib)

The authors led a national project in Denmark on differentiation, in which they sought to strengthen learners' engagement with their own learning and ability to set personal goals in collaboration with teachers. These extracts come from their book *Inspiration til undervisningsdifferentiering [Inspiration for differentiated learning*, published by the Danish education ministry in 1998.

Germany: Being positive about diversity (A. von der Groeben)

Annemarie von der Groeben was for the very experienced Deputy Principal of the Laboratory School, Bielefeld, a school established to promote and evaluate curricular innovation. These extracts from from her 2008 book *Verschiedenheit nutzen: Besser lernen in heterogenen Gruppen [Making use of diversity: Learning better in heterogeneous groups]* published by Cornelsen Press,Berlin.

Section 2: Diagnostic assessment

Synthetic phonics and the phonics check (Margaret M Clark)

Margaret M Clark has many decades of experience as an education researcher, specialising in early education and literacy. The text produced here is an amalgamation of various texts. The data derives from official sources (DfE statistics), although a Freedom of Information demand was needed to discover the effect of pupils' months of birth.

Margaret has recently edited a book about the flawed use of evidence used in promoting synthetic phonics and the phonics check in England and Australia *Reading the evidence: synthetic phonics and literacy learning* (www.witleypress.co.uk)

What could replace the phonics check?(Jonathan Glazzard)

Jonathan Glazzard is Professor of Teacher Education at Leeds Beckett. His research focuses on inclusion, disability, special educational needs and early reading development. This article is adapted from an article originally published in *English in Education* in 2017 (vol 51, issue 1).

Specific references:

United Kingdom Literacy Association (UKLA) (2012) Phonics Screening Check Fails a Generation of Able Readers. https://ukla.org/news/story/phonics_screening_check_fails_a_generation_of_able_readers

Miscue analysis

These two extracts were among many openly available online, and are chosen as examples to illustrate the value of miscue analysis.

EAL assessment framework for schools (Bell Foundation)

The extract reproduced here comes from a substantial set of criteria for all aspects of language and literacy. It was developed by experts at Kings College London and the University of Cambridge, led by Professor Constant Leung, and available to schools free from the Bell Foundation www.bell-foundation.org.uk

Section 3: Supporting teachers in summative assessment

Assessment of primary writing in 2016 (Ros Wilson)

Ros Wilson has over fifty years' experience in education as a teacher, manager and advisor/inspector. She has taught in primary, middle and secondary schools in England and overseas, gaining wide experience in teaching, as well as advising on raising achievement for pupils with English as an additional language and also pupils with special educational needs. She now works as a consultant for Andress Education. She has a master's degree in education. The text presented here is adapted from an article originally published in NATE's *Primary Matters* (2016).

Teaching by numbers: experiences of writing (Nerida Spina)

Nerida Spina is a lecturer at Queensland University of Technology, Australia. Her research interests include the sociology of numbers and the quantification of education, with a focus on the impacts on teachers' work. This article is adapted from one published in *English in Education*, 2017 (vol 51, issue 1).

Assessing A-level English Literature (John Hodgson)

John Hodgson has taught and researched English studies in secondary and higher education in the UK and overseas. He currently edits the journal *English in Education*, and this article is adapted from his article originally published in that journal (2017, vol 51, issue 1).

Assessing primary literacy through grammar (John Hodgson)

John Hodgson has taught and researched English studies in secondary and higher education in the UK and overseas. He currently edits the journal English in Education, and this article is adapted from an editorial for that journal.

Grammar and Great Literature (John Richmond)

John Richmond has been an English teacher in London, an advisory teacher for English, an officer on two national curriculum-development projects in language education, and a local authority adviser. He has been a commissioning editor in educational television. He has published books and articles on English teaching and the role of

language in learning. The extracts reprinted here are from his speech to the 1991 annual conference of the National Association for the Teaching of English.

Section 4: Observation

The Primary Language Record revisited

The Primary Language Record has gone through various revisions and developments since its origins in the UK in the late 1980s, including adoption in various parts of the USA. The text presented here draws on several of them. An early version of the *Handbook for Teachers* (1988) was written by Myra Varrs, Sue Ellis, Hilary Hester and Anne Thomson at the Centre for Language in Primary Education. Recent adaptations are available from www.clpe.org.uk

Assessment through talk (Valerie Coultas)

Valerie Coultas is Senior Lecturer in Teacher Education (English) at Kingston University. She has taught for 20 years in London schools. Her books include *Constructive Talk in Challenging Classrooms* (2006). She has published extensively on spoken language. Much of this article was adapted from a version first published by *Reclaiming Schools*.

Maths is more than the right answer (Gawain Little and colleagues)

Gawain Little is a primary school teacher and maths specialist in Oxfordshire. He is a member of the National Executive of the NUT

(now part of NEU). Jo Horn is a primary school teacher with an interest in RE and philosophy. Steph Gilroy-Lowe has been a home school community link worker and teaching assistant for thirteen years.

The books referred to are:

Jo Boaler (2009) The elephant in the classroom: Helping children learn and love maths. London: Souvenir Press

Susan Lamon (2005) Teaching fractions and ratios for understanding: Essential content knowledge and instructional strategies for teachers. Mahwah NJ: Lawrence Erlbaum.

Jo Boaler (2016) Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages and innovative teaching. San Francisco: Jossey Bass.

Section 5: Portfolios

Assessing primary humanities using portfolios (Tony Eaude)

Tony Eaude was previously a primary school head teacher and now works as an educational research consultant. Dr Eaude has written widely on primary education and co-edited a recent issue of Education 3-13 on the humanities in the primary school. More details of his work are on www.edperspectives.org.uk.

Assessment by portfolio (Kathe Jervis)

Kathe Jervis was involved in many projects concerned with writing and assessment. We draw here on an example from her book *Eyes on the*

child: Three portfolio stories (1996, Teachers College Press)

Portfolios as evidence (Grant Wiggins)

The late Dr Grant Wiggins developed the notion of understanding by design, a curriculum development approach which linked meaningful assessment practice with relevant and deep curriculum knowledges and skills via backward planning. His books include *Educative Assessment* (1998) and *Assessing Student Performance* (1999). His blog can be accessed at https://grantwiggins.wordpress.com/ The extracts reprinted here are from chapter 8 of *Educative Assessment* (Jossey-Bass, 1998).

Portfolios for summative purposes (David Pearson and colleagues)

These examples illustrate the criteria for judging portfolios adopted by a consortium of US states. They appear on p71 of the authors' chapter *Literacy assessment as part of New Standards*, in Coles and Jenkins (eds) *Assessing reading 2: changing practice in classrooms - international perspectives on reading assessment* (1998)

Section 6: Authentic and holistic assessment

Assessing creative learning (Grant Wiggins)

The late Dr GRANT WIGGINS developed the notion of understanding by design, a curriculum development approach which linked meaningful assessment practice with relevant and deep curriculum knowledges and skills via backward planning. His blog can

be accessed at https://grantwiggins.wordpress.com/ The short extracts used here are from his chapter for the *Routledge International Handbook of Creative Learning* (2011, pages 189-91).

A generic rubric for assessing creativity (Grant Wiggins)

The extracts reprinted here can be located on his blog at https://grantwiggins.wordpress.com/2012/02/03/on-assessing-for-creativity-yes-you-can-and-yes-you-should/

Assessing the sixth form 'masterpiece' (Eddie Playfair)

Eddie Playfair is principal of Newham Sixth Form College in East London. He has been involved in campaigns to transform assessment for 16-19 year olds. His website is www.eddieplayfair.com

The Extended Project Qualification (EPQ) (David Leat)

David Leat is Professor of Curriculum Innovation at Newcastle University. This article is adapted from a chapter in his edited book Enquiry and Project Based Learning: Students, School and Society (Routledge, 2017)

Authentic assessment through rich tasks (Queensland: New Basics)

These examples derive from the New Basics project of the State of Queensland, Australia. Unfortunately the website is no longer accessible.

Final comment

The grassroots speak (John Coe, National Association for Primary Education)

John Coe has substantial experience in primary education as a teacher andadvisor. He has been a major influence in the National Association for Primary Education (NAPE) and currently serves as Information Officer and editor of the journal Primary First.