

# Assessment after levels

Ten years ago, two heads of department in contrasting schools presented a powerfully-argued case for resisting the use of level descriptions within their assessment regimes. Influenced both by research into the nature of children's historical thinking and by principles of assessment *for* learning, Sally Burnham and Geraint Brown argued that meaningful assessment could not be achieved by reference to a single measurement scale. Instead, they proposed the use of task-specific mark-schemes that properly acknowledge the interplay between the development of conceptual thinking and substantive knowledge in history. In light of their own continued experimentation and critical evaluation, and drawing insights from other history teachers' research, Brown and Burnham here take readers back again to first principles – the varied purposes of assessment – to help determine the approaches that will best achieve each of them. They share further examples of task-specific mark schemes from across Key Stage 3 (illustrating their conception of students' progress in characterising change and continuity) and provide some essential 'Dos' and 'Don'ts' for history departments reviewing or re-designing their assessment practices.

## Geraint Brown and Sally Burnham

Geraint Brown teaches history at Cottenham Village College (11-18 comprehensive), where he is also Assistant Headteacher. Sally Burnham teaches history at Carre's Grammar School (11-18 selective) in Sleaford, Lincolnshire.

We didn't want to start this article by saying, 'we told you so', but... we were certainly celebrating, as I am sure many history teachers were, when the Department for Education (DFE) announced that 'levels' had been consigned to the dustbin of history:

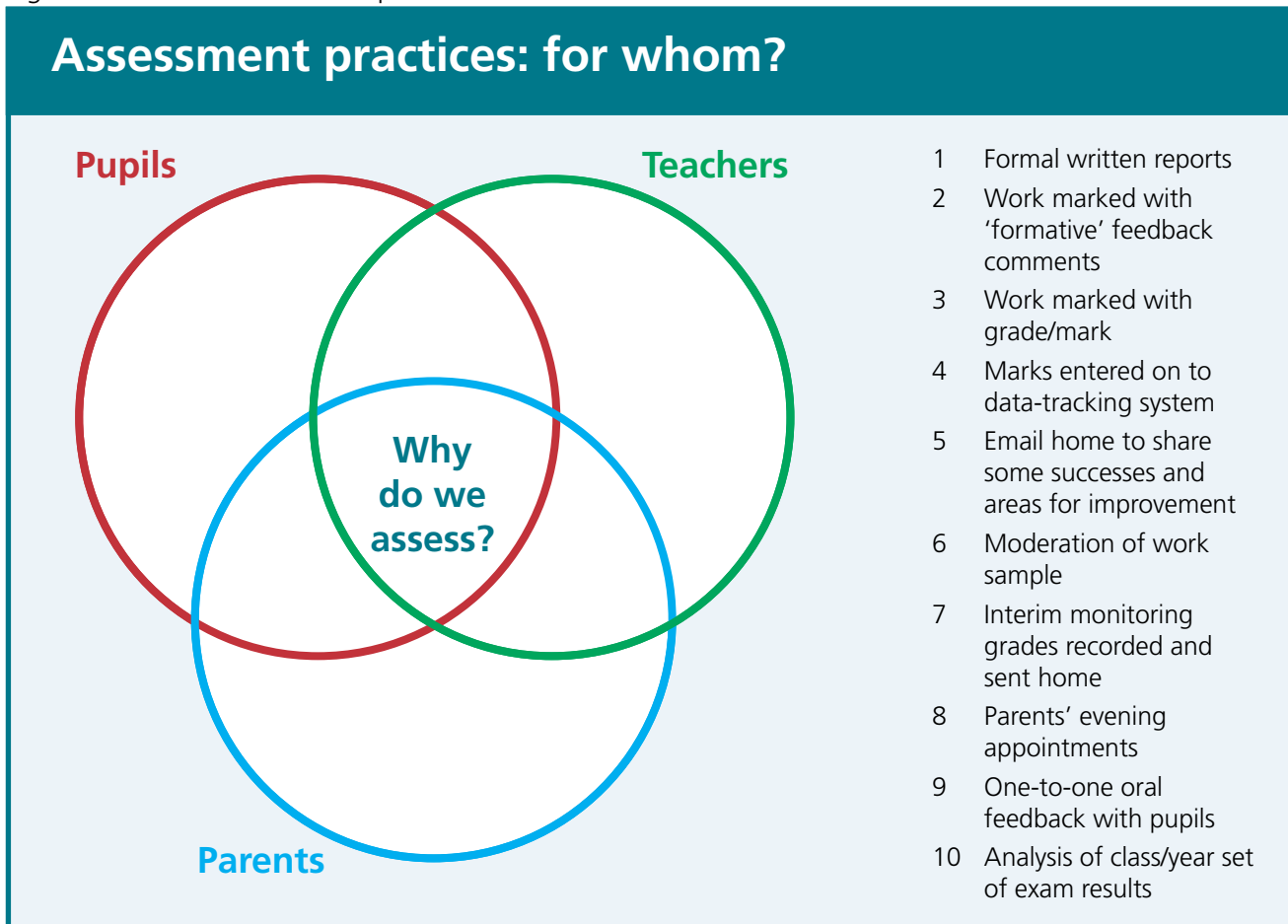
*As part of our reforms to the national curriculum, the current system of 'levels' used to report children's attainment and progress will be removed. It will not be replaced.<sup>1</sup>*

In our previous joint article, written ten years ago, we explored the problems associated with using the National Curriculum Level Descriptions as a means of assessment at Key Stage 3.<sup>2</sup> In our experience, we had found that while level descriptions could be used as best-fit statements at the end of the key stage, they were inadequate for use on a half-termly basis and next to useless for characterising individual pieces of work. While level descriptions had never been intended to be used in this way, this fact was more often than not ignored by senior leaders who seemed desperate for departments not only to apply level descriptions to individual pieces of work, but also to sub-divide each of those descriptions into three further 'sub-levels' (5c, 5b and 5a, for example) so that pupils' progress over increasingly short periods of time could be readily discerned from a spreadsheet.<sup>3</sup> We don't want to repeat ourselves or detail the extensive abuse of level descriptions – most people are only too aware of the problems – but it appears that the abuse is set to continue, at least in the short term, as schools delay decisions about what to replace them with.<sup>4</sup>

When we wrote ten years ago we were determined to focus attention on the purposes of assessment, ensuring that the process was useful to pupils, teachers and parents, as well as serving to provide senior leaders with meaningful and useful data. We also wanted to make sure that assessment wasn't reduced to narrow prescriptions – 'If you define monarchy correctly, you will reach level 5b' or 'If you give three causes, you will reach level 5c' – but rather that the process would give both pupils and teachers a better understanding of how to get better at history. Indeed our aim was to generate further meaningful discussion about what exactly that process involves – a question with which we have not stopped wrestling ever since. In this article, we revisit some of the original principles and practices that we first discussed in 2004 and consider where to go next in light of our own continuing search for the most effective approaches to assessment.

Over the last ten years we have remained adamant that level descriptions are not the way to provide meaningful assessment. While we continued to experiment in our use of task-specific mark-schemes, we also explored other approaches. For example, a pilot group of Year 9 pupils were given copies of the mark-scheme before they completed specific assessment tasks. There is an argument that making the 'success-criteria' as clear as possible to pupils – familiarising them with the standards by which their work will be marked – will enable them to make the best progress and produce the highest-quality work. In fact, we found that such a strategy actually had the reverse effect. Pupils were simply not excited by the assessment task; it was almost as though it had become another tick-box exercise. Rather than the challenge of the enquiry question igniting their enthusiasm and their desire finally to answer

Figure 1: Common assessment practices to consider in relation to their intended audience



a question that they had been investigating for some time, all the pupils' learning seemed to have been washed away by the appearance of the dreaded mark-scheme. Needless to say, we moved away from this strategy by re-focusing on the engaging historical question – with positive results. We also experimented with non-linear mark-schemes, using sets of statements to try to capture the range of thinking that might be required in responding to questions about, say, significance or interpretations. These statements were then used to assess particular responses, allowing us to identify the types or aspects of thinking that the pupils had employed. These groups of statements, however, proved too vague and abstract. While they encouraged a focus on thinking and argument in relation to essential second-order concepts, they completely neglected the development of substantive knowledge and could not adequately accommodate issues relevant to the specific planning and teaching for each individual enquiry. So we returned to the idea that our assessment measures needed to be rooted in a specific task. Time and time again, our experience confirmed that we and the pupils learnt most from using task-specific mark-schemes.

When the demise of level descriptions was announced, we had expected there to be rejoicing in the playgrounds. In fact there has been considerable reticence both about celebrating their departure and about seeking alternatives. Senior leaders are keeping the levels 'for one more year' in the hope that someone will tell them what to do. To some extent, we understand this restraint, given the considerable external pressures and measures that teachers regularly face. However, we believe that now is the time to press forward with further experimentation: to use department meetings

to reignite discussions about what it means for pupils to get better at history; about how to use assessment effectively; and about how to create meaningful information about pupils' attainment and progress. In the sections that follow, we offer our own answers to these questions and share the processes by which we have arrived at them, in the hope that this account of our journey will encourage others to embark on their own.

### Why do we assess?

While it ought to be simple to answer the question, 'Why do we assess?', the system of levels seemed to generate the answer, 'So we can collect data and measure progress' rather than, 'To help pupils get better at history'. Even the DFE's *Assessment Principles* begins its definition of 'effective assessment systems' by stating that they are ones that, 'Give reliable information to parents about how their child, and their child's school, is performing.'<sup>5</sup> The central purpose appears to be to judge *how* pupils are doing, rather than to improve learning; measurement and the comparison of outcomes predominate and the accountability agenda drives policy and practice. The unfortunate consequence, as Biesta has observed, is that 'we end up valuing what is measured, rather than that we engage in measurement of what we value.'<sup>6</sup> In the last ten years, it has therefore felt like the tail has been wagging the dog in terms of assessment in schools. While we do not dispute that it is important to track progress, we stand by the principle that we endorsed in our previous article that assessment is *for* learning. Yet many of the systems that teachers now operate seem largely (if not entirely) divorced from that central purpose. Although the notion that assessment should be *formative* has certainly not

disappeared since we last wrote together, what have clearly changed are the purposes for which assessment has been used and the ways in which it has been deployed. If we are to respond to the challenge and the opportunity presented to us by life after levels, it is time to ask again, 'Why do we assess?' before evaluating and changing *how* we assess.

There exists a great deal of tension between different activities that come under the umbrella of assessment in schools, especially between summative and formative assessment. Consider the list of ten common practices shown in Figure 1 – practices that might typically be conducted (among others) as part of the process of assessment in schools – and think for a moment about why we conduct each of them and for whom: for pupils, for teachers or for parents? There is a place for all these activities if used effectively, yet it quickly becomes clear which have most, and least, value in helping pupils actually to get better at history. Different people in school may well answer the question of why we assess in quite different ways. A data manager may argue that it is to 'monitor and track progress', which is why he or she tends to prefer data presented in an apparently accessible format, whether that is a level (or more commonly a sub-level), a mark or a grade. However, what actually makes assessment *useful* is that it helps pupils, teachers and parents to know and understand precisely what lies behind the superficial score or symbol and what needs to be done to help secure further progress.

### **The purpose of assessment for pupils**

For pupils, therefore, effective assessment and feedback is an essential part of helping them to make progress. Indeed, for pupils, that *is* its purpose. We have not changed our approach in the last ten years with regards to pupils and have continued to give comment-only feedback, to which pupils respond in follow-up activities, both inside and outside the classroom.<sup>7</sup> The only way to give *effective* feedback, in our experience, is by giving precise, diagnostic formative feedback based upon task-specific mark-schemes, as detailed previously and further illustrated below. Summative assessment data for individual pieces of work is recorded, as are holistic judgements about attainment and progress at key points, although this is not attached to the pupils' work and is not shared with them during feedback. Therefore, the focus of assessment is not on helping pupils simply know where they are in terms of attainment scores, but on helping pupils know exactly what to do to get better at history. Black and Wiliam's influential work did much to establish the concept of 'assessment for learning' in schools, and we have continued to adhere to the principles of formative assessment that they set out.<sup>8</sup> Ensuring that the assessment experience is positive and useful for pupils is far more likely with comment-only feedback and where time is given for pupils to respond to that feedback.

### **The purpose of assessment for teachers**

For teachers, assessment helps to establish how, and how far, they are changing pupils' understandings in lessons, thus enabling them to evaluate their teaching and carefully plan for the next steps in the short, medium and long term. This is obviously a very complex process, requiring

in-depth specialist knowledge about history teaching and learning and about particular pupils, which cannot be easily captured in a simplistic overarching assessment framework or numerical data. For teachers to succeed in adapting their teaching effectively, assessment needs to be appropriately related to the specific context. The particular question asked, the wider context of the enquiry, the topic, the lesson and the pupils are all important if assessment is to help teachers work out how to teach better and to take pupils' learning forward. While school leaders will no doubt continue to demand data for tracking progress, such data should not replace the deep and rich knowledge we, as teachers, have of the pupils, and which we use to inform our planning.

### **The purpose of assessment for parents**

In our experience, one of the most alarming changes in the last ten years relates to the way in which summative assessment data has been used with parents. While we were perhaps lucky in being able to resist using levels and sub-levels for individual pieces of work, many schools not only recorded such data but began sharing it 'live' with parents. The problem is that this did not help pupils get better at history because it did not help parents to understand *how* to help their children to do so. The knock-on effect was that conversations via email, over the phone and at parents' evenings began to be about whether a particular pupil was performing at Level 6a or 7c, rather than talking about the pupil's progress in history and what they might do to improve. Thankfully the government eventually recognised this problem, acknowledging in 2013 that one of the main reasons for abolishing the system of levels was because it was, 'complicated and difficult to understand, especially for parents.'<sup>9</sup> Nonetheless, many schools continue to use levels (by which we mean the levels descriptions and not just the numbering system), and are planning to do so for the foreseeable future, at least for whole-school monitoring purposes.<sup>10</sup>

So, as we enter the brave new world of assessment without levels, let us not forget one simple but essential point: assessment is *for* learning and therefore whatever we do next must be planned and evaluated with this principle in mind.

## **Developing Key Stage 3 assessment**

In wrestling with the question, 'how should we assess in a post-levels era?', we have been reading professional literature, talking with our department colleagues, discussing with other departments in school, debating with colleagues from other history departments and talking to our senior leaders. At the moment we are working on a model that includes our end-of-enquiry outcome tasks and shorter knowledge-based tests. Inspired by Riley's ground-breaking article on the power of enquiry questions to shape the curriculum, the outcome tasks that we have devised help us to assess a range of types of historical thinking and have become diverse in nature, including essays, spoken presentations, television documentaries, annotated cartoons and historical narratives.<sup>11</sup> These enquiry-based tasks ensure that assessment is integral to the teaching, bringing together the learning that has taken place rather than being bolted on at the end of a 'topic'. To mark them we use task-specific mark-schemes, which enable us to assess the development of pupils' substantive knowledge

Figure 2a: Task-specific mark-scheme for a Year 7 enquiry relating to change and continuity

	
<b>Excellent</b>	<b>Good</b>
<p>The analysis categorises the <i>types</i> of changes taking place (e.g. in relation to the feudal system and loyalty, law, religion, language) as well as characterising the <i>nature and extent</i> of that change (whether things were switched, uprooted, replaced, reshaped, altered, maintained, etc.).</p> <p>Selects, organises and deploys a wide range of knowledge effectively in order to support their analyses and arguments about change and continuity, perhaps contextualising it beyond the period studied using prior learning</p> <p>By examining how changes after 1066 were experienced by <i>different groups</i> in medieval society (lords, monks, peasants), identifies and explains the co-existence of change and continuity and identifies when things changed, for whom and in what ways.</p> <p>In the context of the work, terms such as 'feudal system' 'religion' and 'law' are used confidently and meaningfully to support explanation and analysis.</p> <p>Reaches a substantiated conclusion about how far England was transformed, which is persuasive. These conclusions are compared and contrasted to the conclusions reached by Schama, which are also explained in outline.</p> <p>A clear argument is conveyed through well-organised paragraphs; the structure is purposefully and deliberately constructed and the written style shows a sense of audience and employs some carefully chosen 'language of change and continuity'.</p>	<p>Accurate and sometimes rich descriptions of changes that took place (e.g. in relation to the feudal system and loyalty, law, religion, language) and some analysis of those changes is offered. For example, may focus on describing and categorising the types of changes or characterising the nature and extent of change/continuity, but probably does not analyse all these different aspects. Justification of the analysis may show some weaknesses.</p> <p>While different groups in medieval society might be mentioned, they may not be linked to specific changes and pupils are unlikely to describe how change and continuity co-existed.</p> <p>Substantive knowledge will be selected according to some discernible criteria, even if they are not explicit. This is organised into a structured account, although the role of the knowledge in supporting explanations may be left implicit and undeveloped. Terms such as 'feudal system' and 'religion' are used when exploring types of change but not always in direct support of analysis.</p> <p>There is an attempt to address the claim that England was 'transformed' but any conclusions are not fully justified. Schama's argument may be described accurately but his claims are not compared explicitly to the pupil's own description and analysis.</p>
<b>Very good</b>	<b>Fair</b>
<p>Direct analysis of the types of change taking place by categorising them, as well as by characterising the nature or extent of change in order to develop simple arguments about change and continuity after 1066.</p> <p>The experiences of different groups in medieval society are described and there is evidence of comparisons being drawn between groups to identify change and continuity happening concurrently.</p> <p>There is a conscious development of the analysis and the account shows evidence of careful, deliberate selection and organisation of information to produce a structure that is directly and explicitly analytic. Terms such as 'feudal system' are used confidently, demonstrating a working understanding of them.</p> <p>There is a reasoned conclusion, effectively linked to the substance of the essay, in which pupils consider how far England was transformed. Schama's argument is described and there is some consideration of how far the pupil agrees or disagrees with his claims.</p>	<p>The response describes some changes that took place or contrasts 'before' and 'after' without explicitly characterising the nature, extent or type of change. Events described (such as feudal system) have relevance but are not used to form clear explanation and analysis. Continuities are likely to be ignored.</p> <p>Some evidence of planning and structure but not used purposefully, for example to examine the type or nature of change. Different groups may be mentioned, although there is no deliberate attempt to compare their experiences.</p> <p>Terms such as 'feudal system' and 'law' are used but there is no evidence that these are securely understood.</p> <p>Reasonable use is made of substantive knowledge, although its selection and deployment may appear random. Some pupils may use everything they have studied whereas others omit key details that could have strengthened their explanations.</p>
	<b>Ungraded</b>
	<p>Does not directly answer the question. Reference is made to the Norman invasion, and to aspects of life in Britain or to key events, but no apparent attempt is made to identify or describe change or continuity.</p>

and their capacity to deploy it effectively and in increasingly sophisticated ways, as well as their ability to think historically.

Another key assessment technique with which we are currently experimenting is the use of short knowledge-based tests. We are beginning to use these on a regular basis to help us to analyse how well pupils are developing their long-term knowledge retention, enabling them to select and deploy the specific information that they will need in answering each

enquiry question and in making comparisons across time and place as their knowledge grows. Some of the questions test factual knowledge – dates of key events, names of key individuals – while others probe pupils' understanding of substantive concepts such as 'imperialism' or 'peasant' or test pupils' ability to construct time-lines from memory or to sequence events. This is a technique that we have always used with GCSE and A-level classes, but it seemed to fall off the radar when teaching Key Stage 3. Since introducing the



knowledge-based tests we have been able to identify those whom Fordham described as the ‘chronologically lost’ much more quickly than when we relied purely on question-and-answer sessions in class.<sup>12</sup> Such regular tests help to ensure that pupils are not left floundering until the final outcome task identifies them as struggling. Pupils have quickly got used to this type of assessment and we have been impressed by the development of their answers in the outcome tasks, as they have become much more adept at using substantive historical knowledge when they have it at their fingertips.

In many ways the assessment strategies that we have been using match Fordham’s ‘mixed constitution’ approach to assessment.<sup>13</sup> Although we have been adapting our approach to fit this principle, we have still had to think hard about how we are assessing each aspect of pupils’ historical knowledge and understanding. The factual recall is relatively easy and ‘marks out of ten’ are uncontroversial in this respect. Writing task-specific mark-schemes is not so easy, however. While we have been using these for ten years now, the fact that we end up re-writing them on a regular basis shows how demanding – but necessary – the process is. Every year, after pupils have completed an outcome task and we have marked it, we look back as a department team to the task itself and to the mark-scheme to see if there are ways to improve it, based on the pupils’ responses. Their work always reveals new things to us about learning in history and sometimes we shift our ideas about what it means to get better as a result. We see assessment as an iterative process, which means we are constantly updating and improving the assessment tasks.

So how do you go about writing a task-specific mark-scheme? We tend to start by considering – again as a team – what we think would constitute a ‘gold standard’ for a particular group of pupils (such as a year group) in answering a particular type of question, in a particular context. By drawing on historical scholarship about that topic and on both professional and research literature about the development of historical thinking, we are able to reach a shared understanding of what progress and attainment should look like and what we need to do in terms of teaching. This often involves considering what we want pupils beyond Key Stage 3 – at GCSE and A-level – to be able to achieve. Such discussions are therefore part of wider conversations about the planning across a key stage and in relation to each individual enquiry and involve rigorous intellectual wrestling with the kind of question being asked and our objectives: what it is we want to see pupils learning *about* and learning to *do* as a result of each particular sequence of lessons. In the following section, we have tried to illustrate this process by sharing our recent experience of developing, improving and assessing some of the enquiries within our Key Stage 3 curriculum that are focused on students’ understanding of change and continuity.

## **Analysing change and continuity: what does it mean to get better?**

The concept of change and continuity has been a focus of considerable recent attention. Like Foster, our department teams felt that we needed to ‘confront our collective confusion’ about this particular concept.<sup>14</sup> Although we may

not yet have a perfect plan for progression across the key stage in relation to this aspect of historical thinking, we feel far less confused than before, having planned new enquiries, developed existing ones, considered what it means for pupils to make progress and found ways to assess that progress.

The series of assessment mark-schemes shown in Figure 2 a-c were developed as part of the planning for individual ‘change and continuity’ enquiries in Years 7, 8 and 9. The sequence of tasks and their associated mark-schemes therefore reflect a planned model of progression in change and continuity enquiries for our pupils. In planning each enquiry and each assessment task, we have considered where and when pupils should revisit similar types of question and have drawn on the work of Foster and Counsell, in particular, to ensure that pupils’ thinking is being moved forward across Key Stage 3 and that we understand exactly what kinds of progress we are looking for in their work.<sup>15</sup> Overall, at different points across the key stage, we wanted pupils to encounter opportunities to analyse change in terms of its extent, nature, type, direction and speed; to appreciate the interplay between change and continuity, and to examine how those processes were experienced by different people in the past. Without explicit planning at enquiry, lesson and activity level to engage pupils in these different kinds of analytical thinking, we knew that they would continue to fail to see change as a process (rather than as an event) and would be stuck at simply organising events into chronological sequences without actually using ideas of change and continuity to make sense of their developing knowledge of events.

So, for example, early in Year 7 pupils tackle the enquiry question ‘Did the Normans transform England?’ – an enquiry which uses extracts from Schama’s *A History of Britain, Volume 1* as a way of engaging pupils with historians’ arguments about how England was changed after 1066.<sup>16</sup> Across the sequence of lessons, pupils focus on characterising the nature of change (whether it was violent or abrupt, for example) and the various types of change (political, social and cultural) and consider the experience of different groups in society (distinguishing between peasants and lords). At the end of the enquiry their task is to write to Schama explaining how far they agree with his argument. As you can see from the mark-scheme (Figure 2a), what we are looking for is whether pupils are able to select and deploy their knowledge in ways that support their descriptions and analyses of the extent, nature, type and scope of change. In considering the experiences of different groups, we are looking to see whether they can identify and explain the co-existence of change and continuity and whether they recognise when things changed, for whom and in what ways.

In Year 8, inspired by Foster’s work, we ask pupils, ‘What kind of reform was the Reformation?’<sup>17</sup> This builds on their work in Year 7 by engaging them in characterising the nature and type of change (religious, social, political). In blending stories studied in depth with overviews across time, our intention is to enable pupils to analyse continuities and to reach conclusions about the balance between change and continuity. We encourage pupils to experiment with different metaphors chosen to help support – and challenge – their thinking about the nature and process of the change. As reflected in the mark-scheme (Figure 2b), pupils are

Figure 2b: Task-specific mark-scheme for a Year 8 enquiry relating to change and continuity

 <b>Year 8: What <i>kind</i> of reform was the Reformation?</b>	
<b>Excellent</b>	<p>A thorough analysis categorises the types of change taking place (religious, political, social) as well as characterising the nature and extent of that change using discernible criteria of their own choosing. There may be recognition that within a particular type of change some changes were more or less 'radical' than others, showing understanding of complex patterns of change. Pupils may discuss the continuity of ideas between the reformers and earlier reform movements (e.g. the Lollards) by drawing on prior learning and showing contextualised understanding. They will show understanding of the way that change affected different people in different ways, and was perceived and construed by different people in different ways.</p> <p>Careful and deliberate selection, organisation and deployment of a wide range of knowledge are used to sustain an argument and explanation about the kind of change that the Reformation represented, leading to a well-substantiated and thought-provoking conclusion.</p> <p>Confident and purposeful use of terms such as 'parliament', 'protestant', 'reform' and 'foreign policy' shows that pupils can use these terms to support their explanations about change and continuity. A range of analytic ideas and language is used to express and provide support for the claims that are made.</p>
<b>Good</b>	<p>Offers some analysis of the <i>types</i> of change taking place, as well as an accurate description of them. Characterises aspects of the nature or extent of change, although the characterisation may not be fully justified (criteria may not be explicit, for example). Shows awareness of the continuity of ideas between the reformers and earlier reform movements (e.g. Lollards) but these may not be fully explained or explicitly analysed.</p> <p>Conscious selection, organisation and deployment of knowledge help to support the analysis and explanation of the kind of change that the Reformation represented, although historical details may be juxtaposed rather than compared or connected. The conclusion will largely be substantiated by and linked to the substance of the essay, though some points may be undeveloped.</p> <p>Terms such as 'parliament', 'protestant', 'reform' and 'foreign policy' are used to support some of the explanations. Some well-judged language may be used to support and express the analysis.</p>
<b>Fair</b>	<p>A description of changes that took place during the Reformation rather than any explicit characterisation of the type/nature/extent of change, although some analysis may be implicit. Where they do seek to characterise change, they are unable to justify their characterisation or the evidence they use to support their characterisation may conflict with it.</p> <p>The range of knowledge and images used may be organised with some sense of logical structure although it may lack evidence of careful and deliberate planning. The conclusion may be thoughtful although it may not be sustained or fully substantiated.</p> <p>Terms such as 'parliament', 'protestant', 'reform' and 'foreign policy' are generally used accurately, but they are not deployed effectively to support the analysis and description.</p>
<b>Ungraded</b>	<p>Does not directly answer the question. Reference is made to key aspects of the Reformation but there is no apparent attempt to identify or describe change or continuity.</p>
<b>Very good</b>	<p>Describes and explains the <i>types</i> of change taking place through categorising and begins to characterise the nature and extent of change using some kind of criteria. They may discuss the continuity of ideas between the reformers and earlier reform movements (e.g. Lollards), drawing on prior learning to contextualise the analysis of change. Analysis reveals knowledge and understanding of some of the complexity of change, such as how different people experienced and perceived change, though this may not be fully justified.</p> <p>The selection, organisation and deployment of a range of knowledge and images is well-considered and will help to support nuanced descriptions and an argument and explanation about the kind of change that the Reformation represented, leading to a substantiated conclusion. This conscious and deliberate exploration of the question is supported by a structure that is directly and explicitly analytic.</p> <p>Terms such as 'parliament', 'protestant', 'reform' and 'foreign policy' are used to support explanations of change and continuity. Different analytic ideas and language are used to express and provide support for the claims that are made.</p>

encouraged to employ new vocabulary to help them express their ideas and to improve the precision of their analysis.<sup>18</sup> In the final outcome task, pupils produce a double-page spread for a textbook, in which they draw together and sum up their analysis, characterising and categorising the nature and types of changes in religious beliefs and practices and making a judgement about the extent of that change.

By the time pupils encounter the enquiry 'How radical were changes in British politics 1800-1928?' in Year 9, they have (we hope!) begun to recognise the types of question being asked about change and may therefore have a greater sense of what is required. This means that they can choose to explore different ways of answering the question. In tackling this enquiry pupils further expand their thinking

about the nature of change and about the interplay between change and continuity by considering the experience of different groups in society, both in terms of class and gender. (See Figure 2c.) There is also a focus on analysing the direction, speed and extent of change by considering how members of those different groups would have viewed particular changes (such as the Reform Acts) in terms of progression, regression or continuity. Pupils evaluate how radical reforms were from the perspective of those who lived at the time and finally from their own perspective, standing back to reflect as historians in the essays they complete at the end of the enquiry.

While all these enquiries are concerned with the processes and patterns of change and continuity, progress is not defined

within and across each enquiry simply in terms of this second-order conceptual thinking. Since the introduction of the revised National Curriculum we have started to re-engage with what it means for pupils to get better at history in terms of their historical knowledge. Although we are only beginning to sketch out exactly what that might mean within these revised mark-schemes, we have considered questions such as whether acquiring historical knowledge simply means knowing more or having a growing ability to use knowledge purposefully through the thoughtful selection and deployment of specific information. We have also asked about the role played by an enhanced understanding and more sophisticated use of substantive concepts such as 'parliament' or 'democracy' in enabling pupils to make connections and comparisons across periods and geographical settings. Although we may not yet have got the balance right – and we are sure there will soon be more to read in the pages of *Teaching History* on the role of knowledge in 'getting better at history', (stimulated in large part by the research and development work reported by Hammond in this issue) – the process of debating such questions as a team has made us think afresh about this issue.<sup>19</sup> It is the freedom offered by the removal of levels that has ignited this debate, since the previous level descriptions made it perfectly possible to ignore the importance of knowledge.

## How should you give feedback and report outcomes?

One of the most worrying trends in recent school practice has been the way in which assessment, recording and reporting have essentially been merged, through a process intended to simplify 'pupil progress data' into a numerical format (or a scale marked out by letters or colours) that means very little, particularly when sent home as a report to parents. Many teachers will probably be under pressure to continue to record outcomes in this way, so that progress can be monitored and under-achievement detected. Despite the good intentions driving this process, there is a danger that subject teams are restricted in trying to design effective assessment practices by whole-school monitoring systems. The question for subject leaders (and one that ought to be asked by curriculum leaders) is how to stay true to the discipline while also providing clear, meaningful information for senior leaders, parents and pupils that is not too simplistic and reductive. Let's forget, for a moment, about the unhelpful practices that others may be seeking to force upon us and consider instead what *would* be appropriate and useful.

The task-specific mark-schemes exemplified in Figure 2 a-c are intended as professional tools to be used by teachers to judge progress and to inform the feedback that we give to pupils. The mark-schemes themselves are not shared with pupils; nor are they simplified or broken down so that pupils can jump through the hoops, since our focus is on teaching and learning history and not on teaching to the test. Written feedback, such as that exemplified in our previous article, is given to pupils for each of the final outcome tasks.<sup>20</sup> This feedback includes annotations on the work to pinpoint specific strengths, or pose thought-provoking questions, as well as a summary of the overall strengths of the work and suggestions as to how the pupil might develop their thinking in future. Feedback such as this

relies on each teacher's knowledge of the individual pupil, of the sequence of lessons and specific enquiry question, and on their understanding of what progress means in terms of substantive knowledge, historical thinking and communication of their understanding. In other words, the teacher is using professional knowledge, informed by the carefully constructed mark-scheme, and operating, 'with a working sense of a gold standard' in order to assess and give feedback effectively.<sup>21</sup> There is, importantly, not a level or a grade in sight. This would immediately distract the pupil from what is important – the feedback with which they need to engage. The approach described is largely the same as that described ten years ago in terms of giving feedback to pupils for 'milestone' pieces of work, i.e. the significant outcomes to each enquiry. Such an assessment policy enables us to provide useful guidance to those actually trying to get better at history so they have a clear sense of the specific features of historical knowledge and thinking that they need to develop and further refine.

Pupils' assessed work, stored in a book or portfolio, obviously contains detailed, useful data about their progress, particularly for teachers and pupils. Such collections of assessed tasks may be less convenient than a spreadsheet, but they are far more useful! Where simple data must be recorded for individual pieces of work, it is easy to assign numerical values to the various 'levels' in the mark-schemes we have exemplified here. These criterion-referenced scores will thus record pupil attainment so that 'progress' from each individual's starting point, as indicated by detailed information about their previous work, can be monitored. The careful wording of each mark-scheme enables us to record something far more meaningful than the previous generic levels could ever hope to provide. Although there is a need for staff to work closely together to create such mark-schemes and to moderate work to ensure that assessment is both valid and reliable, we would argue that this is still more robust and informative than simply using the levels for purposes for which they were not intended.

Figure 2 a-c shows how mark-schemes can be used together to monitor pupil progress in a far more meaningful way. Pupils' responses are assessed in terms of their development and use of substantive knowledge (in relation both to specific periods and topics and to a wider contextual framework) as it is used to create historical narratives, explanations and arguments. Each task involves new knowledge and, even though the second-order focus is the same, it is clear that the final outcome tasks become more analytically demanding. Thus, what constitutes a 'good' response in Year 7 to the question of whether England was transformed as a result of the Norman Conquest is quite different from what is expected as a 'good' response in Year 9 to the question of how radical changes in British politics were in the period 1800-1928. The differences encompass pupils' knowledge, their use of substantive concepts and their abilities to think historically (i.e. their use of second-order concepts) and to communicate their understanding. Where numbers are assigned to these 'levels' for data collection purposes, it is important to note that if a pupil continues to score 7 out of 10 across different pieces of work, according to the planning and the assessment mark-scheme, they will have made huge progress because of the increasing demands represented by

Figure 2c: Task-specific mark-scheme for a Year 9 enquiry relating to change and continuity

 <b>Year 9: How <i>radical</i> were changes in British politics 1800-1928?</b>	
<b>Excellent</b>	<p>Clear recognition of the kind of question being asked and an effective choice of foci to address in developing a response. Effectively conveys a thoughtful argument about the direction, nature or process of change. Exploration of the interplay between change and continuity in relation to different groups in society. Claims about change and continuity are nuanced and patterns or trends in enfranchisement may be considered (progression, regression). They may distinguish between different experiences of change <i>at the time</i> and <i>subsequent</i> analyses of significance of particular developments such as the Great Reform Act.</p> <p>Detailed substantive knowledge is used highly effectively, through measured selection and precise deployment, to characterise the nature/extent of change and support and sustain an argument about how radical particular reforms were. A broad range of knowledge across different periods of study (from enquiries in Y7/Y8) is drawn upon to contextualise the period and strengthen the analysis of change.</p> <p>Use of a range of analytic ideas and language to examine the direction/nature/process of change and to characterise the perceptibility/imperceptibility of change from different perspectives. The argument is conveyed through coherent and meaningful paragraphs leading to an effective and well-substantiated conclusion. The structure has been purposefully and deliberately planned to support the analysis and argument.</p> <p>The written style is mature and fluent. Spelling and use of technical terms is accurate. The response comes across as an original piece of work which engages the reader.</p>
<b>Good</b>	<p>Some analysis is offered of the direction, nature or process of change in British politics, although not all claims are fully explained. Change and continuity are identified for different groups identified, although there is limited analysis of the overall patterns and trends and little consideration as to how they represented regression, progression or continuity.</p> <p>A range of substantive knowledge is used to support explanations and a simple argument is advanced about how 'radical' the reforms were, although this argument may not be fully reflective of the main body of the essay.</p> <p>Some attempt is made to use analytic language to describe and explain the direction, nature or process of change. Different perspectives and experiences may be explained and analysed.</p> <p>The response is organised to produce a structured account, although the relationships between different elements may be left implicit.</p>
<b>Fair</b>	<p>There is a recognisable focus on describing the change/continuity in British politics, through this is not always carried through to attempt an analysis of the direction, nature or process of change. Change and continuity may be identified for different groups, although these differences may be identified in rather simplistic terms and remain unrelated to the overall direction or process of change.</p> <p>Accurate knowledge is used to support descriptions. Claims are offered about how 'radical' the reforms were, although they may not be fully reflective of, or supported in, the main body of the essay.</p> <p>The organisation of the account shows some evidence of planning, but the structure is not used deliberately and purposefully to support the explanation advanced.</p>
<b>Very good</b>	<p>Thoughtful argument and analysis of the direction, nature or process of change in British politics, although this may be under-developed at times. Change and continuity are considered in relation to different groups in society. The overall pattern of change and continuity in relation to enfranchisement may be explored (in terms of progress and regress) though such claims may not always be fully supported.</p> <p>Deliberate and effective use is made of substantive knowledge to support an argument about whether the reforms were 'radical'. Wider knowledge of earlier periods in British history is drawn upon to contextualise change.</p> <p>Analytic language is used to describe and explain the direction/nature/process of change. Different perspectives and experiences are explained and analysed.</p> <p>This conscious and deliberate exploration of the question is supported by a structure that is directly and explicitly analytic, leading to an effective and substantiated conclusion. The written style is fluent, with correct spellings and accurate use of technical terms.</p>
<b>Ungraded</b>	<p>Does not directly answer the question. Mention may be made of voting and of the experience of different groups, but there is no apparent attempt to identify or describe change or continuity.</p>

each successive enquiry and associated mark-scheme. Should their marks drop, they may have produced work of a similar 'standard' to that on a previous task, but this would suggest underachievement or a lack of progress. A higher mark would suggest that, in the context of that enquiry and mark-scheme, the pupil had responded particularly well and made excellent progress. While the mark-schemes and scores show attainment, since they reflect planning for progression across

enquiries (in a year or key stage), the scores will also represent progress in concrete subject-specific, knowledge-specific and discipline-specific terms. In schools where the data manager is obsessed with seeing numbers increasing – and cannot accept that a sustained mark of 7 out of 10 as tasks become more challenging represents appropriate progress – it may be necessary simply to shift each column up to make visible the step up that is actually involved. In such a system pupils



would be awarded more marks – a higher score – reflecting the step up they had achieved in tackling a more demanding task successfully.

Many schools' data tracking systems also involve RAG-rating pupils in terms of progress. This often happens automatically where the data entered (say, a level for a piece of work) is compared with a target based on national expectations of progress from KS2 to KS4. The relevant cell turns a particular colour, with red representing a level 'significantly below target', amber a level 'below target', green 'meeting target' and purple 'exceeding target'. Such a flagging system is not a terrible idea in itself, but it disregards the wealth of knowledge that teachers have about pupils' progress in a sequence of lessons, or over a longer period, that may not be fully portrayed in a final piece of work or represented by the levels. If we were to accept that data systems are simply to 'flag up' possible underachievement we could use it for that purpose and use our judgement to choose the appropriate colour. Pegging it to the numbers entered often means, in our experience, that people choose the level to award a piece of work based on the colour it will turn the relevant cell on the spreadsheet, rather than basing their choice on what the level actually represents in terms of attainment. We would suggest that by using a task-specific assessment scheme, a department would generate more valid and useful data. Separating such task-specific mark-schemes from the flagging system is likely to make both much more reliable when judging how much progress has, or has not, been made.

Parents also need reliable, valid and useful information about pupils' progress. Sharing numerical data is never meaningful *on its own* and cannot help parents understand how to support their children to make progress. In our experience, sending assessment work, even assessment portfolios, home has always proved to be very valuable and is well worth the risk of them not coming back immediately! We tend to do this at fixed points in the year, usually about once a half term, so that parents can read the feedback on specific tasks and therefore understand what the pupil needs to do to improve. Parents are asked to sign the feedback to show that they have read it before it is returned to school. Comments from parents and pupils about this process have been very positive, as parents like to see what their children have been doing and how they are progressing. In addition to communicating with parents at a parents' evening or through a written report, we have found this sharing of the work and of our feedback to be an effective way of helping parents understand the learning in which their child is engaged and how they might help and encourage them to do better. The more we reduce and simplify information about assessment the more likely we are to fail in helping pupils actually to make progress in history.

## Principles and practices to take forward

In light of our experience since drafting our original plea for 'Assessment without levels' and in the spirit of the new freedoms offered by the abolition of levels, we offer the series of principles and warnings set out in Figure 3. We hope that they will inspire and guide other history teachers and heads of department to act boldly in this brave new world. A world without levels is an exciting opportunity. Seize it!

## REFERENCES

- 1 Department for Education (2013) *Assessing without Levels*, London: DFE. Available online at <http://webarchive.nationalarchives.gov.uk/20130904084116/www.education.gov.uk/schools/teachingandlearning/curriculum/nationalcurriculum2014/a00225864/assessing-without-levels>
- 2 Burnham, S. and Brown, G. (2004) 'Assessment without level descriptions' in *Teaching History*, 115, *Assessment without Levels?* Edition, pp. 5-15.
- 3 Ofsted criticised this kind of practice in the 2011 subject report *History for all*, pointing out that in most of the history departments visited by their inspectors 'teachers were using sub-levels in their assessments, usually in response to demands from senior leaders'. Such practice, according to the report, 'was largely unhelpful since the levels were not intended for such minute differentiation or to be used so frequently. They were intended to be used sparingly and holistically to judge several pieces of work at the end of a key stage or, at most, at the end of a year. They were never intended to be used, as inspectors observed in some of the schools visited, to mark individual pieces of work.' Ofsted (2011) *History for all: history in English schools 2007/10*, p. 29.
- 4 For more detail of some of the criticisms see Lee, P. and Shemilt, D., 'A scaffold not a cage: progression and progression models in history' in *Teaching History*, 113, *Creating Progress Edition*, pp. 13-23; Fordham, M. (2013) 'O Brave New World without those levels in't: where now for Key Stage 3 assessment in history?' in *Teaching History*, *Curriculum Evolution Supplement*, pp. 16-23; Ford, A. (2014) 'Setting Us Free? Building meaningful models of progression for a "post-levels" world' in *Teaching History*, 157, *Assessment Edition*, pp. 28-41. Teachers' uncertainty about how they will respond to the new freedoms is revealed by their responses to the Historical Association annual survey. In June/July 2014 around a third of respondents indicated that they were still unsure about how to proceed. Of those who were able to suggest what they would do, responses varied quite evenly between those who intend to continue applying the level descriptors as set out in 2008; those who intend to adapt those 2008 levels in some way; those who will look to GCSE grade systems to devise a related scale and those intending to develop their own measures of progression. See Burn, K. and Harris, R. *Historical Association Survey of History in Schools in England, 2014*, Historical Association, p.1. Available online at: [www.history.org.uk/resources/secondary\\_news\\_2303.html](http://www.history.org.uk/resources/secondary_news_2303.html)
- 5 Department for Education (2014) *Assessment Principles*. Available at: [www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/304602/Assessment\\_Principles.pdf](http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/304602/Assessment_Principles.pdf)
- 6 Biesta, G. (2009) 'Good education in an age of measurement: on the need to reconnect with the question of purpose in education' in *Educational Assessment, Evaluation and Accountability*, 21, no. 1, p.43.
- 7 The use of comment-only marking was explored in detail ten years ago by Simon Butler, drawing on research by Ruth Butler (no relation). See Butler, S. (2004) 'Question: When is a comment not worth the paper it's written on? Answer: when it's accompanied by a Level, grade or mark!' in *Teaching History*, 115, *Assessment Without Levels Edition?*, pp. 37-41; and Butler, R. (1998) 'Enhancing and Undermining Intrinsic Motivation: the effects of task-involving and ego-involving evaluation on interest and performance in *British Journal of Educational Psychology*, 58, no. 1. pp. 1-14.
- 8 Black, P. and William, D. (1998), *Inside the Black Box: raising standards through classroom assessment*, London: Kings College; Black, P. and William, D. (2002) *Working inside the Black Box: assessment for learning in the classroom*, London: Kings College.
- 9 Department for Education (2013), *op.cit.*
- 10 The Historical Association (2014) *Survey of history in schools in England 2014*, London: Historical Association, p. 1. Available online at: [http://www.history.org.uk/resources/secondary\\_news\\_2303.html](http://www.history.org.uk/resources/secondary_news_2303.html)
- 11 Riley, M. (2000) 'Into the Key Stage 3 history garden: choosing and planting your history questions' in *Teaching History*, 99, *Curriculum Planning Edition*, pp. 8-13.
- 12 Fordham, *op. cit.* p.21.
- 13 *Ibid*, p.21.
- 14 Foster, R. (2013) 'The more things change, the more they stay the same: developing students' thinking about change and continuity' in *Teaching History*, 151, *Continuity Edition*, p.8.
- 15 *Ibid*; Foster, R. (2008) 'Speed cameras, dead ends, drivers and diversions: Year 9 use a "road map" to problematise change and continuity' in *Teaching History*, 131, *Assessing Differently Edition*, pp. 4-8; Counsell, C. (2011) 'What do we want students to do with historical change and continuity?' in Davies, I. (ed.) *Debates in History Teaching*, Routledge, pp. 109-123..
- 16 Schama, S. (2000) *A History of Britain - Volume 1: at the edge of the world? 3000 BC-AD 1603*, London: BBC Books.
- 17 Foster *op. cit.*
- 18 This focus on the development of appropriate vocabulary draws on Woodcock's approach to using language to help develop pupils' historical thinking and builds on similar work within enquiries focused on causation. See Woodcock, J. (2005) 'Does the linguistic release the conceptual? Helping Year 10 to improve their causal reasoning' in *Teaching History*, 119, *Language Edition*, pp. 5-14.
- 19 Hammond, K. (2014) The knowledge that "flavours" a claim: towards building and assessing historical knowledge on three scales' in *Teaching History*, 157, pp. 18-24.
- 20 Burnham and Brown, *op. cit.*
- 21 Fordham, *op. cit.* p.17.
- 22 Fordham, *op. cit.*
- 23 [www.cem.org/blog/would-you-let-this-test-into-your-classroom/](http://www.cem.org/blog/would-you-let-this-test-into-your-classroom/)
- 24 Lilly, J., Peacock, A., Shoveller, S. and Struthers, D. (2014) *Beyond Levels: alternative assessment approaches developed by teaching schools*, National College for Teaching and Leadership. Available online at <http://www.psqm.org.uk/docs/beyond-levels-alternative-assessment-approaches-developed-by-teaching-schools.pdf>

*Thanks to Rachel Foster and Matt Stanford, who have contributed enormously through their invaluable discussions with us to many of the ideas and practices presented within this article.*

Figure 3: Dos and don'ts in developing assessment policy and practice



Begin with what you want pupils to learn and then consider how to design assessment systems and practices to reflect this. This will involve planning in the long-, medium- and short-term to ensure that assessment is fully integrated into planning for teaching and learning and may well mean adopting a 'mixed constitution' for assessment across the key stage.<sup>22</sup>

Think deeply, as a department, about progression and reflect critically on current assessment practice. Although challenging, it is this process that distinguishes teaching as a form of professional practice. An unfortunate consequence of the level descriptions, for some, was that it closed down thinking about what progression looks like and how it might be assessed.

Collaborate with other schools and draw on existing good practice, such as that shared on the pages of *Teaching History*, in order to design assessment systems and develop practice. Not only will this help to share the workload, it will challenge and improve your thinking and help ensure that there is a clearer understanding of what 'expected progress' means in history.

Get involved at whole-school level. Don't wait to be told what system you will have to work within and then have to meet its requirements. By offering to help develop school practice, you are far more likely to influence policy in positive ways. Communicate with senior leaders to ensure that they understand what you need from assessment policy as history teachers.

Analyse and evaluate the quality of any new assessment system regularly and rigorously. Consider using (at least some of the questions checklist devised by Professor Robert Coe and shared on his blog to help you evaluate the quality of the assessment you design. It is certainly worth using this list to arm you against any 'weak' externally-imposed structures and systems.<sup>23</sup>



Use the levels as they exist or create something largely similar to the levels. The level descriptions were never intended to be used for formative assessment or individual pieces of work. So, don't try creating a generic linear model of progression that fails to capture the complexity of historical progression and ignores the importance of historical knowledge.

Use GCSE mark-schemes from Key Stage 3 onwards. Such generic mark-schemes that reduce progress to small steps in a simplistic, linear way will simply encourage more teaching to the test. GCSE mark-schemes are weak models of progression that largely ignore substantive knowledge and the complexity of second-order conceptual development, so will not help pupil progress.

Use a single taxonomy (e.g. Bloom's) as a structure for assessment, as suggested by the NCTL's 'Beyond Levels' 2014 research report.<sup>24</sup> Designing assessments and creating displays about making steps from 'description' to 'explanation' and 'analysis' will be meaningless and confusing, particularly out of subject context. It would also be wrong: a rich description characterising a period might be far more complex than a simplistic causal analysis, for example.

Plan your assessment system around external, generic or whole-school structures and systems such as data tracking or league tables. These are not rooted in subject discipline and are too simplistic to be useful. Find a way to make this work after you have the rigour in place.

Use numbers or grades rather than descriptions in an effort to make things easy to do and easy to use. Data has its uses but carefully-crafted descriptions will enable you to capture the complexity of subject-specific progression. Perhaps Einstein had levels in mind when he (allegedly) said, 'Not everything that can be counted counts, and not everything that counts can be counted.'