Exploring the testing ‘effect’ to enable knowledge retention and deployment in the Key Stage 3 History classroom

Nicholas Dennis

A Research & Development Project Submitted for the MSc Learning & Teaching 2016

DEPOSIT AND CONSULTATION OF THESIS

One copy of your dissertation will be deposited in the Department of Education Library via the WebLearn site where it is intended to be available for consultation by all Library users. In order to facilitate this, the following form should be completed which will be inserted in the library copy your dissertation.

Note that some graphs/tables may be removed in order to comply with copyright restrictions.

Surname	Dennis
First Name	Nicholas
Faculty Board	Department of Education
Title of Dissertation	Exploring the testing 'effect' to enable knowledge retention and deployment in the Key Stage 3 History classroom

Declaration by the candidate as author of the dissertation

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.

2. I understand that the Department requires that I shall deposit a copy of my dissertation in the Department of Education Library via the WebLearn site where it shall be available for consultation, and that reproductions of it may be made for other Libraries so that it can be available to those who to consult it elsewhere. I understand that the Library, before allowing my dissertation to be consulted either in the original or in reproduced form, will require each person wishing to consult it to sign a declaration that he or she recognises that the copyright of this thesis belongs to me. I permit limited copying of my dissertation by individuals (no more than 5% or one chapter) for personal research use. No quotation from it and no information derived from it may be published without my prior written consent and I undertake to supply a current address to the Library so this consent can be sought.

3. I agree that my dissertation shall be available for consultation in accordance with paragraph 2 above.

Abstract

This project aims to explore two interrelated issues. First, it seeks to probe recent interest in with research from experimental cognitive psychology regarding the use of low-stakes tests as a learning and memory tool rather than a summative assessment instrument. The idea that frequent testing of key dates, names and information as a way to improve student learning, produce superior performance in examinations and overcome social and economic deprivation has been seen as an exciting opportunity within the history and wider education community in England. By transplanting the basic methodology and placing it into the real world context of a school classroom rather than the tightly controlled laboratory space, it seeks to analyse the performance of students over a term in comparison with a class not subjected to frequent low-stakes testing. Qualitative data in the form of written work and semi-structured interviews are used to probe student conceptions of tests as learning tools.

The second related aspect the project aims to explore is the interest in the teaching of 'knowledge' and increasing the capacity of student memory which has fuelled the desire to explore cognitive psychology as a vehicle to liberate students from limited historical understanding and poor academic performance which leads to poor life chances.[1]

[1] Elements of this dissertation appear in Dennis, N. (2016) 'Cognitive Psychology and testing without guarantees' *Teaching History, Feedback Edition* 164, 22-28

In our collective search to improve learning opportunities for students and their enjoyment of the subject, we all wrestle with particular events, concepts and ideas. One such issue becomes evident in the discussions with students after mock examinations as a history teacher and Deputy Head responsible for academic progress. Before meeting the students, progress is reviewed through tracking data and discussion with teachers and viewing marks gained in normal assessments throughout the year. In the majority of instances, there is nothing in the review to suggest poor performance in the mock examination. This is confirmed by comments from the students who seem to say, 'I revised everything but I didn't really understand the question' or 'I knew my stuff but thought I had written a good answer'. This mismatch between performance at the time of teaching and performance after a few months is particularly frustrating and highlights the key concern driving the research of enabling students not only to develop genuinely secure knowledge (historical or otherwise), but also to deploy that knowledge in an effective and meaningful way some time after the initial teaching and learning phase has taken place. The stakes are high enough when the issue is considered within the context of formal examinations in schools, which is what prompted me to think very carefully about the issue; when the lack of knowledge is positioned as a national problem and articulated by politicians, the consequences are broad and far reaching.

Michael Gove (2009), as Secretary of State for Education from 2010 to 2015 and Nick Gibb (2015), as Minister of State for Schools from 2010 to 2012 articulated their desire to redesign the school curriculum with a focus on 'knowledge', influenced by their reading of E D Hirsch. For Hirsch (1987, 1996, 2006, 2015) the failure of US education, as he saw it, was not directly due to teachers but emanated from the ideas that tended to drive teachers' professional discourse. This discourse, he argued, created conditions in which students were not explicitly

taught, and therefore did not acquire, the necessary background knowledge to make the academic progress expected of them in school. Influenced by the Italian historical materialist, Antonio Gramsci's, criticism of schooling in Italy during the 1920s, Hirsch argued that by increasing 'cultural literacy' – in essence, knowledge – students would be able to access the curriculum and perform at higher levels. Knowledge is regarded not an end in itself, but as a form of *emancipation* which leads to engagement in the economic and political community. According to Hirsch, the best way to ensure that students gain the necessary knowledge and to hold schools to account for student progress is to test the students as tests 'are the only practical way to hold schools accountable for educating all children and are therefore essential to the twin aims of equality and fairness' (2006: 91).

In 2009 Gove had declared that if the Conservatives were to win the next general election they would overhaul the curriculum 'to ensure that the acquisition of knowledge within rigorous subject disciplines is properly valued and cherished' (2009). Although the Conservatives did not win a decisive victory and instead could only form a coalition government with the Liberal Democrats, Gove did as he had promised and sought to reform the history curriculum along his stated line. The first draft, published in February 2013, caused a great deal of uproar (Evans, 2013; Burn, 2015), in part at least, it was argued, because the document did not show any recognition that the history education community had already been working towards helping students acquire knowledge within a subject discipline (Lyndon-Cohen, 2013).

The idea of knowledge improving life chances and academic performance is undeniably attractive. Spurred on the prevailing political mood, Hirsch's ideas and his use of cognitive

psychology in justifying his claims 'proved formative in the curricular and pedagogical theorising of a (predominantly) young generation of teachers who sought to advance a 'knowledge-rich' curriculum' (Fordham, 2015a: 56). In particular, Daisy Christodolou (2013) and Joe Kirby (2013a), English teachers and former 'graduates' of teaching charity Teach First, were early supporters of Hirsch's ideas through various blog posts and conferences such as ResearchEd. They both attempted to transplant Hirsch's culturally and historically specific analysis of US education to England. In doing so, they also sought to replicate elements of Hirsch's analysis which used cognitive psychology to show the enfeebling effect that teacher professional discourse had on classroom practice. The danger, as Hall (1996) pointed out when using the work of Antonio Gramsci in his analysis of Britain in the Thatcher years, is that using the work of someone who is not stepped in the environment means that it should be '…delicately disinterred from their concrete and specific historical embeddedness and transplanted to new soil with considerable care and patience.' (Hall, 1996: 413). The transposition of the US centric analysis infused with cultural touchpoints of Manifest Destiny and exceptionalism ignores the long development of educational thinking in England with influences from antiquity, religion and intellectual movements including the Renaissance and the Enlightenment (Lawton & Gordon, 2002).[2] This 'forgetting' made it easier to merge two distinct beliefs together. The first was that cognitive science shows us what makes effective instruction (Kirby, 2013b; Christodolou, 2014; Radice, 2016). The second was that low-stakes testing aided retention, and deployment, of knowledge (Didau, 2013, 2015; Fletcher-Wood,

[2] There is a question regarding Hirsch's reading of Rousseau and his 'Romanticism'. In the work Considerations on the Government of Poland, Rousseau argued that it is 'education that must give souls a national formation, and direct their opinions and tastes in such a way that they will be patriotic by inclination, by passion, by necessity.' (Rousseau, 1986: 176) This is not a 'romantic' gesture where individuality and natural development are free.

2013). The merging of these ideas allowed a particular view to be expressed. If testing served to increase knowledge, by not using them in your teaching, you were contributing to the poor progress your students were making and thus to a bleak national picture.

Framed in this way, it is not hard to see how these two ideas combined and why it held an attraction in the initial work of Part 2. However, observation of the discourse around testing as a *guaranteed* way to improve learning, and considering the reality of assessment in the history classroom and throughout the school, led to a reconsideration of the claims made by the evangelists who were not history subject specialists. If the effects of testing were as robust as the discourse suggested, then a thorough investigation within the context of a history classroom and would hopefully resolve the recurring problem described above and lead to improved performance.

For the purposes of the dissertation, assessment is viewed as an integral part of education in that it helps us to see where students are in their learning. Harlen and James (1997) distinguish between formative and summative assessment in that formative assessment is positive, not criterion referenced and covers a number of instances where skills and ideas are used. In contrast, summative assessment is linked to reporting cycles, is referenced against public criteria and it provides some form of quality assurance (1997: 372-373). There is of course an overlap between these forms and this is embodied in the use of tests. Traditionally considered to be a form of summative assessment, tests usually refer to a form of measuring a person's knowledge of a topic, knowledge domain or skill and the outcome is usually reported or has element of high-stakes in that results can lead to set changes/class changes for students. What is clear is that there is some confusion about the use of these terms by

teachers (Black, 2015) and there is a confusion about how they are used in classrooms (as discussed below) and this is also wrapped up in the political discussion where testing is seen as the mechanism to raise standards (Harlen & Deakin Crick, 2003: 170).

The lens through which this exploration of assessment, testing and the influence of cognitive psychology will take place is via the discourse of history teachers and educators in England which Fordham (2016) suggests grew out of dealing with curriculum issues and 'drew upon a highly subject-specific knowledge base' (Fordham, 2016: 135). The notion of discourse used in the project is extended to include the Historical Association (HA), the journal *Teaching History* which is published by the HA, the Schools History Project and project CHATA. Within this discourse, there is an agreement about what history teachers seek to assess. Although there are a variety of ways to characterise what is taught in history classrooms, it is usually considered to be constituted by two types of knowledge as Ashby and Lee explain:

> Substantive history is the content of history, what history is "about." Concepts like *peasant, friar,* and *president,* particulars like the *Battle of Hastings, the French Revolution and* the *Civil Rights Movement,* and individuals like *Abraham Lincoln*, *Marie Curie* and *Mahatma Gandhi* are part of the substance of history. Concepts like historical *evidence, explanation, change*, and *accounts* are ideas that provide our understanding of history as a discipline or form of knowledge. They are not what history is "about," but they shape the way we go about doing history. 2000: 199

The interplay between substantive and procedural knowledge demands careful consideration when placed within the context of assessment and there was clear anxiety surrounding the content, knowledge and assessment practices involved in the introduction of the National Curriculum (NC) of 1991 (Aldrich, 1991). Soon after its introduction, there was a recognition of the difficulty in integrating assessment of substantive knowledge with procedural knowledge in a national system. By trying to create a format that would work across a variety of schools, the substantive knowledge of history had to be specified. As this approach was

rejected, it created a limited focus on procedural knowledge as the framing device for assessment. The NC was reviewed in 1995 to provide more clarity regarding assessment by simplifying the targets to help provide a consistent structure to teacher assessment. The intention was that these refined targets would help teachers grade student work at the *end* of each Key Stage. The reality for teachers was markedly different. Schools, now having to report examination data in league table format from 1995, began to track the performance of students over the course of Year 10 and 11 to ensure positive examination results. The idea of one, end of Key Stage 3 assessment, as advocated by the NC was markedly different form the process of frequent, regular monitoring of students' progress that schools were adopting with exam classes.

To ensure that students were ready for post-14 qualifications, schools began to use the descriptors as a way of tracking performance through individual pieces of work. The effect, as Burnham and Brown (2004) detail, was problematic for history teachers and students. For teachers, the key aim of assessment is to help students get better at history and that assessment should be used for learning (2004: 5). However, the descriptors were limited in the sense that they did not capture the interplay of substantive and procedural knowledge. Alongside this, they did not account for the 'changing ideas, patterns of reasoning and layers of knowledge that make up *progression* in historical learning, and they certainly do not offer guidance for designing the detailed *learning paths* that will secure such progression' (2004: 5).

Assessment for Burnham and Brown should be based on professional thinking about progression in history including knowledge *and thinking* and should be marked against task-

specific mark-schemes as the interplay between substantive and procedural could be captured because the substantive aspect would be determined by the school. Burnham and Brown suggest that a central aspect of any assessment practice is to change pupils' understanding in a lesson or series of lessons and to help the teacher adjust and improve their teaching. Their exploration of the term 'imperialism' is an instructive:

> Just defining imperialism – as though it had one fixed meaning to be learned – is not the point…[i]n reality, the way a pupil uses a demanding abstract noun will continually change as knowledge broadens and new questions are raised. Our job is to get pupils to carry on reflecting upon it, to allow its meaning to change throughout their time learning history in school and beyond, not to learn a single fixed definition and have done' 2004: 8-9.

Substantive knowledge is important but historical thinking is more than a fixed concept and any assessment instrument must attempt to capture this complexity if it is to help students and teachers because ''getting better' often cannot be seen as finding the next increment. It is about revisiting, reshaping, reconnecting (2004: 11).

The discourse on historical learning was influenced by the work of Black and Wiliam (1998a) who in their research on assessment tools defined assessment as an activity used by teachers and students, which provides information to be used as feedback to adapt the teaching and learning activities they are engaged in (1998a: 2). For Black and Wiliam, assessment became formative when the evidence produced is used to adapt teaching as opposed to summative assessment when the evidence was used as a judgement or a reporting point. Ideally, the assessment would be formative, providing feedback to the student and teacher about the next moves to enhance learning.

Kitson, Husbands and Steward (2011), reinforce the belief that the best assessment systems are embedded within realistic contexts. For history teachers, written and spoken assessments

can be difficult formats for students to show their understanding of the past. Whilst identifying the correct date of an event or putting a group of events into the right chronological order is important, activities such as these demonstrate little valid evidence of pupils historical understanding (Kitson et al., 2011: 97). As a result, assessments should be complex enough to help capture the notion that while a student recalling the date of the Battle of Waterloo is useful, it does not generate a complete sense of the pupil's historical understanding. Factual recall knowledge in this way is necessary but far from sufficient (Kitson et al., 2011: 97)

The challenge, as Burnham and Brown point out is to make it workable in the classroom. This leads to a tension to achieve clarity in assessment criteria without reducing the marking of the work to atomised chunks, because learning rarely presents itself as such. Where it has been placed within clear criteria, as Burnham and Brown point out, it becomes very difficult to make a clear decision on where it stands in relation to the substantive knowledge and procedural knowledge that comprise historical thinking. This can be seen at GCSE and A Level where the prescribed curriculum with specified content obscures the capacity for a student to understand where they are and what a teacher needs to do next. The focus then becomes on the grade achieved or reported and this is the problem identified by Burnham and Brown in their conversations with their Senior Management Teams (SMT) at Key Stage 3, but without recognition of the wider pressure of the examination system further up the school. High stakes testing at GCSE and A Level 'has shaped teacher practice to 'teach to the test', and to adopt teaching styles which meet the needs of testing: an emphasis on short-term recall, and marks and grades' (Kitson et al., 2011: 101).

Resolving the tension between the wider societal demand for measuring schools based on performance and the demands of assessing for historical learning can be seen in the notion of 'getting better' or 'progression' in history. Lee and Shemilt (2004) argue that progress can occur for any history student in terms of making notes, giving presentations or writing essays. Being 'better' is not generic; it links to a particular activity. For example, 'better' essays will be attuned to the question and have a clear structure:

> 'The kind of achievement at stake impacts on how we measure progress. Progress in the amount of information students are able to remember is often measured quantitively. Thus, in public examinations in history in Year 11, ticks used to be given for items recalled, subject only to fairly weak criteria of relevance; the more ticks the better. We may be unhappy with the simplicities involved in this, but we can hardly dispute that it involved a kind of progress. 2004: 13

However, progression is something different from progress:

> Progression was juxtaposed with 'aggregation' to emphasise that progress in history could be more than an increase in the amount of information pupils could recall: learning history was not just learning 'one damn thing after another'…it arose from dissatisfaction with the aggregationist assumptions that seemed to be implicit in examinations, widespread among curriculum managers in schools and enshrined in classroom practice. 2004: 13

For Lee and Shemilt, pupils did not just add to their information of the past; they acquired understandings that changed as they considered the subjects they studied more deeply. The change was also affected by individual circumstances and across topics. Until history educators could readily understand how the concepts of history are linked and decoupled, 'we cannot indulge in simplistic syncretism and lump everything together' (2004: 15) through a simple assessment task that produced a grade to be recorded.

From this brief survey, the discourse on history teaching seems to be ambivalent on assessment *systems* due to the difficulty involved in historical thinking. Assessment is seen as necessary to help teaching planning, move students on (in a variety of ways) and to fulfil professional responsibilities through report and tracking. It is also seen as something to be

wrestled with as a concept, in teaching, and especially in the reporting of grades for tracking and measurement systems.

The measurement involved in external testing at GCSE and A Level have a powerful backwash effect on earlier stages of the curriculum as suggested by Luff (2016) which in turn creates the 'myth of constant linear progress for every child' (2016: 38). This effect should not be underestimated or limited to the curriculum as examination thinking is present in league tables produced by leading newspapers, parental choices regarding where they live and the school they choose and articulated clearly through teacher professional development and appraisal (Elliot, 1991: 93-117). In this sense, the terrain that the history education community works within when thinking about assessment is defined by historical, political and social forces beyond its control and is similar to Gramsci's characterisation of ideological hegemony. For Gramsci, ideological hegemony was not just a mechanism that one ruling and dominant group uses to coerce a lesser group to bend to their will. It was also a system of domination that impacts on the beliefs, explanations and values related to education that also explains the existing social setting as beneficial for everyone through the production of a 'common sense' that is used to justify actions. In other terms, the ideological framework of the society 'manufactured consent' (Herman and Chomsky, 1988). Viewed in this manner, the reporting of grades on individual pieces of work is seen as 'common sense': a way of ensuring, and assuring, the 'quality' of education to a variety of interested parties. It is this 'common sense' approach to assessment that is present in the discourse of government and schools and seen in the struggle that Burnham and Brown have with their leadership teams.

If Gramsci provides a way to think through the problems faced by the history education community when considering assessment and historical learning, his work also provides insight into the practices of history teachers and educators identified in the literature above. For Gramsci, you could resist ideological hegemony by adopting a 'war of position' or a 'war of manoeuvre'. A 'war of manoeuvre' involves physically resisting the focus of domination. An alternative strategy to dealing with a relative stable social situation is by adopting a 'war of position' by building up of alternative ways of thinking and creating alternative resources within the existing social structure (Gramsci, 1971: 229-235). Gramsci's work has been used by Giroux to theorise the role of the teacher as a 'transformative intellectual' (Giroux, 1988: 125) as teachers are engaged intellectuals in the sense that any human activity cannot be abstracted from the functioning of the mind. Moreover, teachers play a role in the shaping the conditions and purposes of schooling through curriculum design, pedagogy and indeed, action research. As such, the discourse on historical learning and how assessment should work by thinking about 'progression' rather than progress, and historical learning with substantive and 'second order' concepts, can be read as the provision of alternative intellectual resources to resist the pressures of SMT and the temptation to relapse into 'examination' thinking. Trying to maintain this balance means adopting a difficult position in which it is understood that although exam grades are important within the existing system, there is a clear understanding that most of what is of value in the study of the past, cannot be captured by an exam grade. (Lee and Shemilt, 2004: 22).

In this sense, the literature above suggests that there is a 'thick' understanding of assessment in the history education community, one that is in tension with a 'thin' or 'aggregationist' view of assessment manifested in the 'common sense' of report grades, progress measures

and attainment targets demanded by the communities beyond the history department. The 'thick' notion of assessment outlined above is intimately tied to the difficulty involved in historical learning and the work of Burnham and Brown, Kitson et al., and Lee and Shemilt all provide alternative intellectual frameworks for thinking about assessment.

In one sense, a focus on testing seems to be in opposition to the 'thick' view of assessment demonstrated by the history education community but the literature of testing from cognitive psychology seeks to move beyond the test as a summative assessment instrument. Roediger and Karpicke's (2006) article suggests that a memory test has a dual purpose. In line with summative assessment, a test can be used to assess what someone knows but they argue that it enhances *retention* of the material. They conducted two experiments to explore the hypothesis. Participants studied two short passages of prose covering general science. In the first experiment, the initial phase of the experiment consisted of four seven minute periods with those in the 'study' condition being able to go over the material twice by using two of the seven minute periods to read the one of the science texts twice. Participants in the 'study/ test' phase could only study the material once and were tested in the subsequent period. A final test took place five minutes, two days or a week later. Where the final test was five minutes after the initial experimental phase, participants who were able to study the material twice performed better than the students who studied and were then tested. The performance of the groups changed when the final test was taken two days or a week later from the initial learning phase. Participants in the study/test condition performed better (2006: 250). In their second experiment, the focus was on repeated studying and repeated testing. Participants either studied the material four times, studied it three times and then took a recall test or studied the material once and took three tests. Then they either took a

final test 5 minutes after the initial learning phase or a week later. Mirroring the first experiment, those who had studied the material four times outperformed all other participants in the test taken after five minutes. In the test a week later, the situation reversed with the participants taking more tests in the initial study phase recalling more of the information (2006: 253).

Roediger and Karpicke go on to discuss several theoretical approaches that may help to explain their results including the view that practising the format of the final test in the initial learning phase helps retention and that testing helps increase the number of 'retrieval routes' used in memory which helped recall (2006: 254). The also collect participant opinions on the predicted success of the study methods with repeated studying seen as inflating the confidence of students.

What is particularly interesting is the notion that an instrument normally used for summative assessment purposes promotes retention of knowledge even after a poor performance. This process of increased retention after a delay is called the 'testing effect' and it is incredibly enticing, mostly because testing, it seems, serves as formative assessment without instructions (to the student or from the teacher). However, there is a clear environmental difference in that the study is based in a lab and not in a classroom. There are other examples where the 'testing effect' seems to to hold in college classrooms (Bangert-Drowns, Kulik, Kulik & Morgan 1991) but the makeup of the participants differs dramatically from a classroom with younger students who are not academically successful enough to reach college-level education. There is also the format to be considered: students in history classrooms in England are generally tested on material they have learned *over a period of time*. The novel

material presented in the experiment and the relatively short learning phase are markedly different to the conditions in schools where a single lesson may be longer than the initial learning condition in the Roediger and Karpicke study. There is also the question of domain relevance in that it is not linked to the discourse around historical learning with its interaction between substantive and procedural knowledge but is focused on *general* substantive knowledge.

This was a question considered by researchers and as Carpenter, Pashler and Cepeda (2009) suggested,

> [n]early everything we know about the testing effect has been derived from the way adult college students learn information in relatively well-controlled environments – most often, in the laboratory. We would be on stronger ground in advocating widespread use of testing at all levels of education if we had stronger evidence showing their efficacy within more diverse educational settings. 2009: 761

Using 8th Grade students (aged 14-15), Carpenter et al., conducted a series of experiments based on what students were taught in a unit of work over a series of weeks. After they had finished the taught unit, the students were asked to review the content either one week after the study unit had finished, sixteen weeks after the unit had finished or not at all. The review sessions consisted of testing what had been learned and gaining immediate feedback on the test, while others re-read the material. After nine months, all students were tested again and the results mirrored the other research on the 'testing effect'. The students who had taken a test after they had finished the unit of work (either a week after or sixteen weeks after) performed better on the final test than the students who had studied the material again a week or sixteen weeks after the initial learning phase. Both the study and testing groups also performed better than the students who had no review at all (2009: 765-766).

The adaptation of the testing effect experiment to a classroom setting and working with younger students manages to build on the work of Roediger and Karpicke and addresses the classroom limitation. By placing the testing research within the context of a more socially and economically diverse group of participants, the possible problem of homogeneity inherent in the pool of college students tested was removed. The format also exceeded the time constraints in the initial Roediger and Karpicke work to better reflect the organisation of school curriculum and when assessment occurs. Despite the appearance of a resolution regarding the 'testing effect' and younger students, the conclusion that 'testing was significantly beneficial to 8th grade students' retention of facts' (2009: 767) requires close scrutiny, especially when compared to the statement that 'the absolute levels of retention after 9 months in the present study were very low. Thus, it seems possible that the absolute benefits observed in these data are very small' (2009: 767). In the immediate review group, the difference between the no-review group, the study group and the test/study group was around a single mark. This was mirrored in the delayed group. The maximum average mark achieved by any student on the 15 items tested was around 4 marks.

Considering that some external examination grade boundaries may span three marks, Carpenter et al's work may be useful if an increased score for recounting facts was included in the GCSE and A Level history assessment which is not the case. However, it is worth returning to the notion of 'progression' used by the history education community and what constitutes historical knowledge. The tests, and their outcomes, equate learning with recall of basic information. This is, in the words of Lee and Shemilt, progress, but not 'progression'. It does not cover the multiple meanings involved in the use, and more importantly, the understanding a student possesses when they use the term. The focus on the substantive

knowledge of history is interesting yet it ignores the complex interplay between the substantive and procedural knowledge even in the prescribed GCSE and A Level courses. As a result, testing, despite its formative aspects, displays a 'thin' view of assessment. Knowledge can be aggregated, improved and measured by a series of tests and on this basis, it seems to fit very clearly within the 'common sense' view of assessment that the history education community has sought to usurp.

A complex formulation of assessment is important when you are able to resist the pressures of 'common sense' but it is incredibly difficult when your students and your school can only succeed by placing a 'common sense' approach to assessment in the high-stakes world of terminal exams. Encouraged by English teacher David Didau's blog post (2013) on the 'testing effect' as a way to help students retain and recall knowledge, Donaghy (2014) experimented with a series of low-stakes tests with his history students. His aim was to improve their 'contextual knowledge' after poor performances for which the majority of students achieved below their target grade for a particular piece of work. Donaghy carefully identified specific knowledge items such as dates, names and people included within the topic being studied. Ten items of knowledge were originally identified for the weekly tests and students were tested on the same items over a number of week until they had gained 'mastery' (near full marks). The reported results showed that students did improve their scores on the same test after a four-week period and the use of contextual knowledge also improved in their responses to GCSE questions in their end of unit assessments. (2014: 47).

The reported improvement seemed impressive because students were able to produce an effective response to questions that demanded substantive and procedural historical

knowledge to succeed. Testing, when used by an history educator was formative, appeared to facilitate an interplay of substantive and procedural knowledge *and* thus apparently showed 'progression'.

The increase in the weekly test scores in a school history education setting mirrors the main findings on testing from cognitive psychology and seems to provide further compelling evidence that 'testing effect' works. The assumption is reinforced by Donaghy's argument when he links testing to social advancement:

> I am increasingly convinced that the various 'gaps' we hear much about in education, and the link between demographics and destiny that my school aims to break, are to a great extent rooted in a knowledge gap which facilitates the educational 'Matthew Effect', where the knowledge-rich get richer and the knowledge-poor get poorer. If we are to confront this effect head on, then we need to lose some of the collective squeamishness we hold as a profession about the learning of 'mere facts', rote methods of doing so and repeated low-stakes testing of pupils throughout secondary school. (2014: 51)

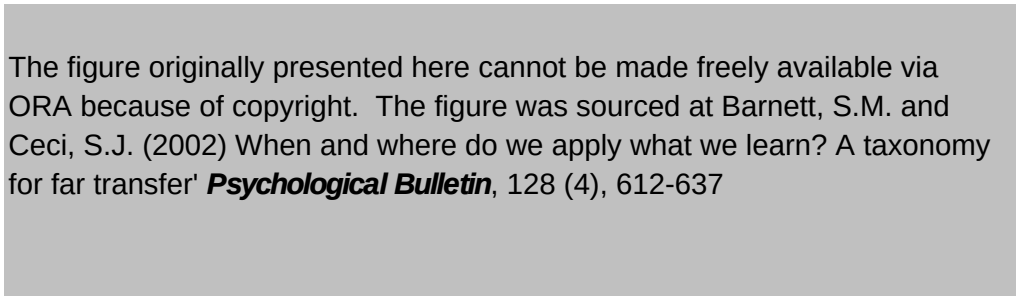
This concluding statement shows the powerful 'backwash' effect of high-stakes testing outlined in the discourse of history teachers and a clear link to the wider social forces beyond the school.

Donaghy's approach to assessment is encouraging and attractive as it seems to resolve the tension between historical learning and the 'common sense' framework that assessment is articulated within. It also seems to address the issue of time and content in that it does reflect the normal teaching cycle in schools and the end of unit assessment is more than a recall of substantive knowledge.

It seems to balance the 'thin' and 'thick' notions of assessment outlined above. Moreover, despite the caveat that it is impossible to say that low-stakes testing caused this improvement, the article is based on the assumption that it does and that testing should be extended to other teaching as it will similarly lead to an improvement in students' contextual knowledge and in their exam performance. It does, however, share two main assumptions with the other literature on testing in that it assumes that the technique is applicable to all students and that testing will increase student performance in a variety of contexts.

The literature on testing assumes that once the retention of knowledge has been secured, that knowledge will be transferred and available to use in a variety of contexts and prove durable over extended periods of time. This assumption masks an ongoing debate on what exactly is meant when the phrase 'transfer' is used and led Barnett and Ceci (2002) to devise a helpful taxonomy of transfer to help provide some clarity to the discussion.

Figure 1 The taxonomy for 'far transfer' devised by Barnet and Ceci (2002: 621)



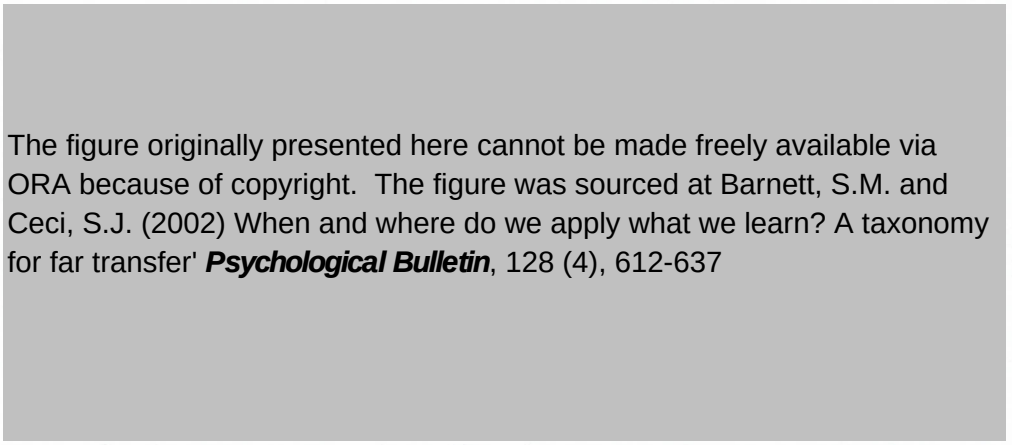
Applying the taxonomy provides fresh insight into the claims of cognitive psychology and Donaghy's research. In relation to the 'content' part of the taxonomy, it seems reasonable to assume that Donaghy was asking his students to make a performance change and to recall key information and recognise the demands of the question set. Using the 'context' aspect, and focusing on the 'temporal' and 'modality' strands, there is a lack of clarity around the timescales used by Donaghy but it is reasonable to locate the end-of- unit assessments that he used in the third of the temporal columns (i.e. 'weeks later') and in the second column of the modality strand (i.e. 'both written, multiple choice v. essay'). Roediger and Karpicke's research when viewed through the lens of the taxonomy also indicates more 'near' transfer on the temporal and modality strands. Carpenter et al.'s work compares well on the temporal strand yet when placed within the modality aspect, its 'transfer' is limited. In all three examples of the literature reviewed, the fact that learning occurs does not mean that it will necessarily translate to a new setting in all contexts yet this is an assumption that is made.

Methods

Exploring the 'testing effect' with students in history lessons fits within the frame of action research which is defined by Elliot (1991) as 'the Study of a social situation with a view to improving the quality of action within it' (1991: 69). The notion that understanding is not enough and that a change should occur is similar to Freire's (1996) critique of education as a counter-hegemonic strategy to dismantle to the existing social order (Cunningham, 2014: 402). Whilst not going as far as Freire in 'realising' humanity from oppression, the action research model is appropriate to explore the claims made about the benefits of the 'testing effect' for education and helping students retain information as part of the dominant political discourse exemplified by Michael Gove's policy intentions. The action research model also provides the means to test the theoretical and experimental claims within a particular classroom setting (Elliot, 1991: 47), addressing the theory/practice problem. To this end, the project adopts a 'critical' theoretical perspective in that it stands apart from the social context, the normative discourse on testing and its advocates asks if it is the panacea it is purported to be in relation to history education (Hammersley, 2012; Cox, 1981: 129).

Following Cohen and Manion's (1989) definition that action research is a 'small-scale intervention in the functioning of the real world and a close examination of such an intervention' (1989: 217), the project was limited to two Year 8 history classes and would cover a term. As a result of the Part 2 feedback and discussion on how to gain rigorous results, a request was made to the colleague in charge of the timetable so that performance between classes could be compared. This would enable the exploration to move beyond the single class example used in Donaghy's research. Due to the time constraints involved, as a new post had been secured for the end of the autumn term, a simple action research cycle was adopted based on Altrichter, Posch and Somekh (1993: 207)

Figure 2 The action research cycle from Altrichter, Posch and Somekh (1993: 207).



The practical theory was that testing would increase knowledge. Scores from the weekly low-stakes tests were collected, as were the scores from assessments and homework tasks. The low-stakes tests mostly centred on substantive historical knowledge in order to keep in step with the published literature. The common assessment tasks were marked using a mark scheme that had been constructed by other colleagues and had been used for a few years in the department. Assessment data was also scrutinised and moderated by the Head of Department to ensure consistency of marking across all staff. A 'mixed methods' approach was used as it would 'generate a more accurate understanding of social phenomena' rather than using a single approach (Biesta, 2012: 147). Using an action research approach also meant that there was direct relevancy to the regular discussion and review of educational assessment throughout the school and the day-to-day work of teachers.

Interviewing students provided an insight into what testing and assessment in general looks and feels like from their point of view. Interviews 'give access to other people's perceptions,

including crucially the thoughts, attitudes and opinions that lie behind their behaviour' (Altrichter et al., 1993: 101). A semi-structured interview provided the best means to ask particular questions about tests, preparation for examinations, and history teaching, while also allowing for digression if it occurred during the normal flow of the conversation. The concern with interviews being biased towards the person constructing the questions was noted and feedback from the supervisor ensured that they were open ended. The power situation involved when teachers interview students was also recognised (Cohen and Manion, 1980: 317-318) and students we reassured before and during the interviews that there were no correct answers to the questions and the aim of the interview was to understand their thoughts.

The research was conducted and data collection occurred during the Michaelmas term (September to December) 2015. The need to focus on small groups within the classes meant a process of selection. To mitigate any potential bias in selecting students, each student was assigned a number and a member of the support staff was asked to pick six numbers from each class set. Once the students had been identified by this method, their Middle Years Information System (MidYis) baseline scores were used to ensure an even spread of scores within in each class and a similar banding spread between classes in order to prevent any potential skew in the results. All student work was marked in accordance with department and School policy and copies were submitted via Google Classroom to help retention and to ensure a complete record was kept for reflection during the process and afterwards. Examining the written work students produced would provide an important source of data for the project as it not only mirrors the usual form of assessment at Key Stage 3 but it would also make it easier to analyse similar work over the three different, but related pieces. The

written work would also make comparisons between classes simple. Initial discussions with the Head of Department about the project in the academic year before the research was conducted made it clear that although the scheme of work for the Tudors could be adapted slightly, any research work had to remain within its general framework so all students could complete the common assessment tasks that were used for tracking and monitoring but also attempt the examination at the end of the year.

The initial interviews were conducted in the first few weeks of the autumn term and in the last few weeks. All interviews were completed during the school day during Tutor/House time of no more than thirty minutes or during the extended lunchtime before sport. Students were allowed to select a time to suit them and they were told that they did not have to take part if they did not want to. The interviews were recorded with a digital audio recorder and then listened to on the day of recording to ensure the data was captured and to gain initial impressions and an overview. Due to the fact that two external presentations based on the research had been booked in the terms that followed, all audio files were listened to again and partial transcription occurred, which reduced time (Altrichter, Posch and Somekh, 1993: 97). Eventually all interviews were transcribed and coded using NVivo software.

Following Saldaña (2009), the written work was coded in at least two deductive cycles. The first cycle of coding allowed for the creation of a heuristic to think more carefully about the data for the second cycle as 'coding is only the initial step toward an even more rigorous and evocative analysis and interpretation' (2009: 8). The second cycle of coding allowed for the creation of categories, sub categories and further codes. As the work was analysed and digested further, a decision was made to replicate the structure of historical knowledge with

an overarching focus on substantive and procedural knowledge with sub-categories. The coding of the interviews was approached in an inductive manner to allow themes to arise from the conversations. A general heuristic was used initially and then refined when the decision was made to focus on a subsample of interviews.

As the work for the research project was part of the normal way of working as a teacher, it was subject to the Modus Operandi for the MSc (Learning and Teaching), which allows for research to be conducted within the context of the School. The project gained approval from the University of Oxford's Departmental Research Ethics Committee in August 2015, conforms to BERA guidelines, and as such the Principal granted permission for the research to be conducted. Both classes were informed that I would be using their work as part of a research project, that the same content and resources would be used, and that I would be trying out some teaching ideas to see if they were useful or not. In this sense, they were all made aware that a research project was underway and their work would be utilised. Because the claim of low-stakes testing improving recall and performance had not been validated, the possible advantage to the class taking the regular low-stakes tests was not deemed an ethical problem. Moreover, if the intervention proved to be successful, then it would have been used across the institution. All students in both classes and in the sample groups were told that their work and any interviews would remain anonymous.

Findings

Building on the work for the Part 2 project and to examine whether regular low-stakes testing would help students retain and deploy knowledge effectively over a long period of time, a replication of the tests discussed in the literature review would be used. The major difference

was that instead of using the same questions every week, the questions would be mixed to match the scheme of work. The questions were all presented on a PowerPoint slide and the students had five minutes to write down their answers. After the allotted time their responses were collected and the correct answers were then displayed on the board. After quickly going through the responses, the lesson carried on as normal.

Figure 3 Results from the low-stakes tests.

Student	Test 1	Test	Test 3	Test 4	Test 5
1	7	9	9	10	5
2	4	5	7	10	7
3	6	5	7	8	5
4	5	5	7	8	6
5	4	3	4	7	3
6	4	2	4	6	5

The initial results were congruent with the literature in that students seemed to be able to retain more information in successive weeks as a result of testing. Tests 3 and 4 were very similar and the results showed the largest increase in performance by two or three marks for three students. The final test was on material that had just been covered in lessons and saw a decrease in performance from test 4. This came about after reflecting on the weekly tests and considering if they were now too easy because the test items have been seen a few times.

Just before the half-term break, students in both classes had to complete the following essay outlined in the scheme of work: 'Why did Henry break from Rome?' The essays were marked according to the department mark scheme and the results from the sample groups and the average marks for each class can be seen in Figure 4 below.

Figure 4 Results from the first essay

Mark achieved for an essay on why Henry broke away from Rome	Average mark achieved by students in the focus group in each class (out of 20)	Average mark across the whole class
Class subject to regular testing	11	13.86
Class not subject to testing	6.4	12.86

The first feature of the essays coded in the testing class was the reference to substantive historical concepts/objects in their work, such as 'relics', 'indulgences', 'Purgatory' and 'ceasaropapism' to help explain why Henry decided to break away from Rome. For example, student 6 wrote:

> *People called for a change to the church because they were abusing their powers. They were abusing their powers by selling fake relics. For example, in 1535 Cromwell and his men found enough pieces of the cross to make a whole cross which was very unlikely.*

Student 3 stated:

> *Martin Luther criticised the church for selling Indulgences (ways of cutting down your time in purgatory), and especially fake ones … They also sold fake relics.*

In the example above, student 3 explains what the term means and further explanation was also employed when discussing the term 'ceasropapism'. This term was deliberately taught and used in the tests because it normally appears in A Level Tudor studies and not at Key Stage 3. Student 1 wrote:

> *Another part of New Thinking is Caesaro-Papism, which is where the King can control the land of England, but also the Religious side of England, which links with having his own protestant church.*

Student 6:

> *…Henry made himself the leader of the Church, and he took lots of land; This was called Caesaro-Papism.*

The small sample of students in the testing class seemed general comfortable deploying these terms in their essays. The non-testing class sample also referenced 'ceasropapism' in two instances but the majority of responses seemed to centre around the bible being translated into English and did not cover 'new thinking', of which ceasropapism was linked with the humanism of Erasmus. The testing class produced around 50% more coded references to substantive historical concepts in their writing in comparison with the non-testing class.

Names were also included in the low-stakes tests and students in both classes referenced Catherine of Aragon and Anne Boleyn. Student 2 from the testing class mentioned a wider variety of names including Henry's mistress Elizabeth Blount, which did not appear in any other answer in the samples across classes. No student in the non-testing class mentioned Erasmus, supporting the coding of concepts above where students in the non-testing class did not consider humanist thought in relation to Henry's break with Rome.

The essays in the testing class used dates and events more liberally in the work to fix events in their writing. Student 6 in the testing class wrote:

> *On the 31st of October 1517 a monk called Martin Luther came to the church.*

Student 3

> *In 1509 he married Catherine of Aragon, his brother's widow*

Student 4

> *William Tyndale was not only a scholar but in 1521 also became a priest.*

Not all the dates used were correct. For example, some students confused the birth of Edward in 1537 with the birth of Elizabeth 1533. However, fewer dates were used in the non-testing class sample, and of the four dates coded from this group one student in the non-testing class (Student 4) referenced one date twice:

> *In 1534 the Pope announced that Henry's marriage to Anne Boleyn was invalid.*
>
> *…so in 1534 a source says that 'the act of supremacy made the king head of the English church not the pope'.*

In comparison, five of the essays in the testing class sample referenced dates to fix arguments.

Historical assessment is more than substantive history, and following Hammond's (2014) work the essays were also coded for procedural knowledge examples. One important feature of Hammond's analysis of second order concepts in her students writing was that they were aware that they were in an argument and that their claims could be contested (2014: 33). For the students in both classes, where awareness did exist, it took the form of a hierarchy of reasons. Student 3 in the non-testing class wrote:

> *One of the main reasons and probably the best reason was for love and for son*
>
> *The second most important reason in my eyes…*

Student 1 in the non-testing class stated:

> *The most important reason why Henry broke away from the church was the fact that he wanted love and a son*

They reiterated this argument in the concluding line of the paragraph:

> *This is the most important reason to why Henry broke away from Rome because the church in Rome were the only people stopping him from getting his divorce, marrying Anne and having a child that Henry longed for.*

Causal language, such as 'trigger', 'exacerbate' and 'provoke' (Woodcock, 2011) were absent from all answers. This was not a particularly surprising result, as this was the first essay the classes had produced in the year and very little work had focused on causation at the start of the year.

Coding revealed that students in the focus group from the testing class used explanations more in relation to substantive knowledge or in advancing their argument. For example, when discussing the abuse of church authority, one student in the testing class wrote:

> *This is the third reason why Henry wanted to break away from the church because with this he could get the English public to stand behind him and would make his case stronger due to the fact that he could use these things to go against the pope saying his churches weren't doing religious things and weren't doing their proper duties all of the time.*

In comparison, a student from the non-testing class stated:

> *My points are important because this is a definite reason to why Henry would close the Catholic Church down because these crimes and disobeying the rules were not the correct path to follow if you want to make it into heaven.*

The majority of students in the testing classes used to dates to fix certain events in their writing, demonstrating a good level of chronological understanding in their writing. Out of the six essays viewed, five made use of dates to fix events in their work.

In comparison, only two of the essays in the non-testing class mentioned dates and there were only three instances that were correct. Where another date was mentioned, it was used incorrectly:

> *And in 1537 Anne was pregnant and Henry had to marry her otherwise it would not be his baby which he really wanted.*

For the common assessment, the following question was set: 'The most important reason that Henry VIII broke away from Rome was because he wanted a divorce'. How far do you agree with this statement?' The framing of the question was deliberate as the department wanted the students to use a form of argument to make their case.

All of the essays made references to concepts and names. Whereas the name Christopher St German appeared linked to Ceasropapism in many of the pieces, for Student 1 in the testing class, this link had disappeared:

> *An idea that someone had told him about and spoken a lot about was Caesaro-Papism, where the king would rule over the land, but also control the religious side to the country.*

The difference between classes in terms of names cited in the first piece of work was reduced and they both became limited in scope. The other names associated with 'new thinking' such as William Tyndale and Martin Luther were not used by any of the non-testing class, building on the lack of citation for Erasmus in the first essay. For two students in the testing class sample these names did appear, reflecting their consideration of the issue in the written work. Student 3 wrote:

> *…Henry was also having trouble with the churches, and people like William Tyndale and Martin Luther were thinking anew and coming up with ideas that began to justify and influence Henry's thinking…*

Fewer mentions of relics and indulgences appeared in the task and the difference between the two classes in terms of the concepts used did reduce. It was the best essay of the group (which gained full marks according to the mark scheme) that mentioned purgatory. What was clear, taking the lack of people identified into account, was that the majority of students in the sample appeared to narrow their use of factors in their writing in comparison with the previous task.

Dates were used by all students in the testing class sample but were used in different ways. Student 4 of the testing class wrote as the final line to their work:

> *Henry was desperate for a son and heir made him break away from Rome in 1536.*

Student 1 from the testing class sample used the same date to fix visitations to monasteries and but seemed to confuse this event with the start of Cromwell's investigation:

> *In 1536, Henry ordered his 'Vicar General' Thomas Cromwell to send some commissioners to monasteries and look for some of the bad parts of the church…*

For the non-testing class sample, the dates used were mostly limited to 1533 as a way of fixing Henry's divorce and his marriage to Anne, with no mention of the chronology in the short-term events between 1533 and 1536. The one other example of a date being used referred to Henry marrying Catherine in 1509.

Returning to the use of procedural knowledge as a key element of historical assessment, there were more examples of being aware that they were in an argument but the strength of this varied. For example, Student 1 in the testing class sample wrote:

> *There are many other reasons why Henry wanted to break away from Rome, but the main reason, in my opinion, was to get a divorce so, he could annul his marriage with Catherine of Aragon, marry Anne Boleyn, then he would be able to have a Baby in a marriage that wasn't cursed so his family could carry on, in their dynasty.*

The use of 'in my opinion' recognises that this claim could be challenged and this appeared at the end of the piece. For some of other students in this sample, awareness of being in an argument started and ended with the first line. Student 6 said:

> *I fully agree with this statement…*

Student 4 wrote:

> *I do not completely agree with the statement 'The most important reason that Henry VIII broke from Rome was because he wanted a divorce' as I personally feel that it shared equal importance with influences from the scholars Erasmus, Christopher St German and William Tyndale.*

Student 4's use of the phrase '…I personally feel' is interesting because it seems to indicate a general view but is followed by a justification for considering the influence of individuals challenging the orthodox thinking about the Church.

For the non-testing class sample, awareness of being in an argument also varied but had improved from their first essay. Student 4 in the non-testing class made it clear that they agreed with the statement in the question and created a hierarchy in their response:

This was the first and the most important reason why Henry wanted to break from Rome.

The final question all students were asked to complete was a source task asking them to use their own knowledge alongside the source to explain why a report on the Fountains Abbey in 1536 was sent to Henry. Student 1 was able to relate the source to their own knowledge by referring to the date and Henry asking Thomas Cromwell to and explain the answer:

The report in Source A was sent to the king, Henry VIII in 1536 because it was 1536 where Henry demanded Thomas Cromwell to send commissioners to all of the monasteries in England and Wales…[t]he report was sent to the king to help him control the monasteries, gain wealth, land and power.

Student 2 made no direct reference to the source but was able to explain eventually that the report was used to close down the monasteries. They also used their own knowledge again to refer to parliament's passing of the act in 1536, which closed the smaller monasteries:

Source A is an example that would give Henry the perfect excuse to close down the monasteries and nunneries. this report is among the many which were sent to parliament and in their shock passed an Act in 1536.

Dates were not present in Student 3's answer but they were able to identify a number of other religious houses that were visited by Cromwell's investigators:

The Fountains Abbey was one of the worst, but other monasteries such as The Crossed Friars Monastery, London, Woolsthorpe, Lincolnshire, St Edmund's monastery…

The response from Student 4 outlined that not only was the information provided to justify Henry's action in breaking with Rome, it was added to the fact:

that they were selling fake relics and indulgences which Henry had also found out about gave him a perfectly good reason for leaving the church.

What was novel in their explanation was they recognised that the monasteries played a vital role in the life of the common people by providing 'shelter and food' and the reports helped to justify why they should be closed.

Student 5 explained:

The report in source A was sent to the king in 1536 because he was looking for reasons so he could catch out the Abbot, which would make Henry on the upper hand.

The student was not able to link the source or the date to the issue of the Break with Rome and the Dissolution of the Monasteries, but took the report as a way for Henry to gain superiority for unknown purposes. Examples from the source were used to bolster their argument that they were doing terrible things but they were not able to place the events in England with the wider historical event.

Student 6's response was interesting because not only did it demonstrate the purpose of the source, it also included a number of dates before 1536, mentioned Cromwell and referenced the source itself. It is worth displaying the whole response:

This letter was sent by one of Henry's commissioners. The King would have sent the commissioner to spy on the Abbot of Fountains to give him evidence to bring down another monastery. In 1536 the start of the dissolution of the smaller monasteries started. Henry didn't like the monasteries because when people needed help for example they went to the Abbey and Henry felt as though the Abbot's had more power than he did which obviously made him very angry. So in 1534 Henry made himself the Supreme leader of the church and made Thomas Cromwell the day-to-day controller of the church. When he made himself the Supreme leader he started to take down monasteries and by 1540 there were no monasteries left in England. In the source it says 'the Abbot wasted his woodland, he kept 6 mistresses and 6 days before the

commissioner's arrival he committed theft.' So this letter was sent from the commissioner to the king as evidence to get rid of another monastery.

Student 6 seems to have a very strong grasp of the chronological events around the period, demonstrated by stating that the smaller religious houses were closed, linking Henry's leadership of the church and noting Cromwell's appointment as Vicar General in 1534.

Student 1 from the non-testing class produced an answer that explained the purpose of the source clearly, mentioned key dates, and included details such as:

…the monasteries were cheating the people, by selling fake relics, and selling 'indulgences' bits of paper that were supposedly signed by the pope. This report, along with the others made at the time would give Henry a perfect excuse to dissolve the monasteries.

Student 2 in the non-testing class response explained the purpose clearly, mentioned Thomas Cromwell's role in the dissolution, used examples from the source and mentioned other religious houses being closed:

The report was sent to the king in 1536 because that was when Thomas Cromwell was looking for reasons to shut down the church…

Figure 5 Results from the common assessment

Marks achieved for both pieces of common assessment	**Average mark achieved by students in the focus group in each class (out of 15)**	**Average mark across the whole class**
Class subject to regular testing	11	10.61
Class not subject to testing	6.4	8.52

At the end of the term, a new role had been secured in another school and the class were taught by another teacher for two terms. All students then sat the same exam paper six

months after being taught the topic. The Tudor section of the paper asked students to explain the message of a pictorial source 'The Family of Henry VIII' and to use it, along with extracts from the Act of Supremacy and from a speech made by Henry VIII, to answer the following question: '"*Henry VIII broke from Rome because of his desire for a male heir". How far do the sources support this statement?*' Students were instructed in the paper to use '*details from the source AND your own knowledge to help you explain your answer*'. Any chance of bias in marking was mitigated by the fact that the all the papers were marked by another teacher and moderated by the Head of Department.

Access to papers was limited due to no longer being employed by the school yet the marks students gained and the sight of the some of the papers revealed some interesting insights. The longer response question was similar to the question that students had previously completed in their longer common assessment piece. The focus group within the testing class again performed better than the focus group in the non-testing class, but the gap was not as wide. However, when students in the testing class completed a task that demanded procedural and substantive knowledge that they had not been previously been tested on but were *taught*, the performance gap was much narrower. The most interesting result was achieved by comparing the *average* mark for the two classes across the whole of the section on the Tudors, where the gap was now almost non-existent. Examining individual responses revealed a complexity not accounted for in the testing literature. In one particularly striking case, it was clear that the student had a strong grasp of the knowledge needed for the task, but he did not consider the sources at all which reduced his overall score. Other students in the testing class were able to combine elements of their own knowledge with points arising from the sources, but they were not very precise in the knowledge that they brought to bear.

Figure 6 Results from end of year exam.

	Average mark (out of ten) achieved on the longer response question by students in the sample group in each class	**Average mark (out of five) achieved on the source response question by students in the sample group in each class**	**Average class mark (out of 20) for the Tudor section of the exam paper**	**Average % mark across the whole paper**
Class subject to regular testing	6	2.6	8.42	59.2
Class not subject to testing	4.6	2.4	8.38	58.2

Fourteen students were originally interviewed for the research project. One student did not attend the final interview. Due to the large amount of data and transcription, a subsample of three from the testing class and the non-testing class were selected for coding and analysis.

Student 1 in the testing class subsample seemed to have a complex and self-reflective idea of history in that had a relational understanding of past and present that had developed during the term:

> *I think we study it because we're looking at the past, of the happenings in England and other countries in the world, and it's very useful to know.*

They stated in the second interview:

> *We learn history so we can understand what's happened in the past, and we can learn culture, and learn more, like how we're here, in some ways.*

Student 6 in the testing class also displayed a nuanced understanding about why history should be studied in school that developed over time. In the first interview they said:

> *... when we learn about things that people have done in the past, such as mistakes or big achievements, we can learn from them, or carry on doing what they did, and kind of learn from them, I suppose.*

In the second interview they explained that we study history:

> *…for different reasons. I think, personally, mostly to learn from it, because I think, for example, World War II, some of the things that people did was really not acceptable, and I think we learn from that. We know now that it, kind of, causes great pain and stuff.*

Student 4 stated that young people studied history because it was 'fun' and that it helped to build a sense of community:

> *Like, last year, in Year Seven, we learned about xxxxxxxx Castle, and it's nice to find out how the town or school you're in, and its heritage, and how it was formed.*

In the non-testing class, a similar situation was found in that history, for Student 1 was,

> *…about the country's past. So we know about our heritage.*

This rather nuanced view of why history displayed by the students in the subsamples was set aside when the students were asked about the types of things they needed to know in history. Dates, names and a focus on monarchs seemed to be identified as the key aspects, rather

than concepts and procedural knowledge. The focus for most of the students in both interviews was the 'what' of history, not the 'how' or 'why'.

When asked to recall what was studied the previous year, topics were recalled simply or not with any real sense of a link. Sometimes the content was relayed in a confusing manner or according to personal preference with very little explanation.

> *Last year we studied about William the Conqueror, and the motte-and-bailey castles, and why they built them, to conquer land*

Student 2 in the non-testing class wrote:

> *Battle of Hastings. The Battle of Hastings was basically one big war to get hold of a throne. So, there was William of Normandy… There were three other people. One was a Viking and I can't remember the two others. Or, one was from Hungary…*

In the answers regarding whether the subject was useful, the focus seemed to shift towards summative assessments as they moved into the year groups above.

> *Yes it is if you want to go forward with GCSE history*
>
> *Yes, because if we didn't study… Well, if I just decided to forget it all, then in my GCSEs or something or in another exam if I've forgotten it all then there's a high possibility of me failing the exam.*

Student 4 in the testing class believed that 'history could help you in the future' but then moved back to focus on monarchs and how it would help in the 'upper years', perhaps indicating an exam focus. Only Student 4 in the non-testing class suggested that it was useful to pass something on:

> *Yes, because when you're older, you can tell your kids that what you've learnt helped, and tell them everything you know, what you learnt when you were younger.*

On being asked what the term 'knowledge' means to them, the students in both subsamples generally equated knowledge to mean facts:

> *…knowing the facts, a certain amount of facts.*
>
> *How much you know, and if you know a question or not.*
>
> *Like how clever you are on something. Like, what you know. So, if I know that one plus one equals two, then someone could say that is knowledge.*

Extended into the question about the importance of factual knowledge, all the students in the first interview seemed to suggest that history had more in common with a *science*, rather than an art, as interpretation and debate were not important. If you did not know your facts, then you did not really know history. For Student 6 in the testing class, the explanation of the importance of factual knowledge in history is revealing:

> *…you have things like Classics, which are kind of stories that have been jumbled with, and not one hundred per cent fact. So I think history is a kind of complete different subject, in a way, because it is fact, and yes, it definitely happened.*

The juxtaposition with Classics as a 'story' and history being a 'different' kind of subject because there appeared to be no 'story' revealed that the students held a view of history as almost scientific. This was reinforced when they were asked about what other kinds of knowledge was needed in history. Student 4 stated that they did not know any other forms of knowledge that they needed to know in both interviews but seemed to be struggling with procedural knowledge as an aspect of history:

> *…you could have a source which was, like, from then on, which you… You have a better chance of it being factual and actually true, other than a historian who did it nowadays, because they're often just going by someone else's report. So, it helps, factual knowledge.*

The students recounted a variety of techniques when asked what was the best way to learn something including self-testing, which seemed to underscore their understanding of history

as ‘fact'. The use of flash-cards was described but it was normally related to either an end of year exam or a test. Self-testing strategies seemed to be a resource that was used only when it was necessary to perform in some form of assessment, which was usually summative.

The students in both classes thought tests were useful for a wide variety of reasons. Tests were perceived as being useful for teachers due to the fact that they provided a way of measuring how much had been learned by the student, or how much ‘progress’ a a student has made. In the non-testing class, students expressed negative feelings towards tests:

> *I probably don’t like them…*
>
> *…they’re putting the pressure on you, so then you try to get them right…*

This was not reflected in the answers provided by the subsample in the testing class. Student 6 in the testing class said in the second interview:

> *…everyone says that they don’t like tests, but after all, it helps you learn from your mistakes. Also, I think when you have a test, it really helps to refresh your mind. I think when you have a test and you revise for it properly, it stays in your mind for longer. Yes.*

Student 4 of the testing class stated:

> *They were useful, yes, because every week, like, I looked at it, then I was like, “I can’t remember that name. I’ve got to get it next week.” (Laughter) I did sometimes, but then sometimes I got it right, and the next week, forgot it. So, yes.*

Students in the testing class were asked a number of additional questions about their experience of the low-stakes test they had uses. During their use in class, students voiced that it was better to call them *quizzes* rather than tests and Student 6 asked for clarification when asked how useful were the tests:

> *The mini quizzes on Tuesday?*

They continued:

I found them useful, because I remember the first one I took was really difficult, and I don't think I answered many questions, really. Then I found because we kept doing them every week, and it was a bit like flashcards. It's just really quick little questions, dates and names, and whatnot. After a while, it just becomes stuck in your mind. As soon as you see the question, you know the answer, and you can just write it down. It becomes just really easy, and it's really helpful, because it's going to be the same in the test.

The students in the subsample had a positive view of the quizzes and this was reflected in their answers on whether the tests mattered:

Well, they mattered to me because I knew exactly what was going on, and I knew what I'd done and hadn't done right. Then, in a way, they didn't really matter because they wouldn't go towards a grade, or they wouldn't really threaten us. They wouldn't make us move sets or anything. It's just to get some ground and know what we're doing.

In keeping with the literature on regular testing, it was clear that low-stakes testing can help to secure and promote student recall of key information in historical answers that are similar to the original tests and that are taken relatively quickly after the initial teaching and testing. The increasing scores of the students in the testing class week after week seems to show that the 'testing effect' works in the weekly tests because it replicates a 'near transfer' situation on an number of strands on Barnett and Ceci's taxonomy of transfer.

Figure 7 The taxonomy of transfer highlighted to reflect the use of low-stakes testing in the School

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at Barnett, S.M. and Ceci, S.J. (2002) When and where do we apply what we learn? A taxonomy for far transfer' ***Psychological Bulletin***, 128 (4), 612-637

The effect seems to continue into the first essay where the testing class student sample were able to include more names, dates, use a variety of concepts and have a generally more secure understanding of chronology than the sample of students from the non-testing class. In all of the other written responses, the testing class and its sample outperformed the non-testing class based on the marks achieved and the amount of substantive knowledge coded in the responses. Even here, when compared to Barnett and Ceci's taxonomy, transfer of learning is still 'near' and is similar to Donaghy's (2014) reported increase in references to concepts, names, dates and events in his student's end of unit source questions. What is of interest is the final class average on the first essay where the gap is one mark in favour of the testing class and the sample group scores in the respective classes. When the smaller sample groups are compared, the data suggests that testing has a substantial positive outcome due to the difference of almost five marks between the samples. However, the class group averages seem to indicate that the gap between the testing and non-testing class was small (one mark), albeit positive. There appears to be a positive effect for using low-stakes tests but the extent of the performance gap reported in the literature was not present when viewing the data from the perspective of class comparisons.

Historical knowledge, however, is much more complex than linking the study of history to 'facts' and 'dates' makes it difficult for students to 'do history' with its link to procedural knowledge (Lee, 2005: 33-34). For the students in the testing and the non-testing class, the

use, and evidence of, procedural thinking was relatively low in terms of being aware that they were in an argument (Hammond, 2014) and using the language of causation. The interview responses from the subsample students revealed that they viewed history to be fairly mechanistic and scientific due to its key constitutive element of 'fact'. This may partially explain why there were so few instances of procedural knowledge in their responses but in examining the scheme of work, it is also clear that this is not a focus for the department at the stage of the year apart from a few lessons. Another explanation seems plausible when considering Elwood and Murphy's (2015) idea that assessment systems act as 'cultural scripts'. For Elwood and Murphy, learners and assessment systems are not autonomous entities that merely interact with each other. This way of thinking about assessment is a legacy of psychometric testing which would have us believe that learners acquire knowledge which is in turn fixed and then can be measured. Instead, assessment systems should be understood, and explained by, the social dimensions forming the system and the effects the system has on societies and individuals (2015: 186-187). For the students in the testing class, the low-stakes tests referred to substantive historical knowledge. Combined with a scheme of work which privileged substantive knowledge, it is unsurprising that both elements helped to shape and consolidate students' views on the meaning of knowledge in history and also constitute their written work to the detriment of procedural knowledge.

Taking this idea further, recognising that the use of the tests is more than just a transaction, an action resulting in a consequence of testing secures knowledge, we can explore the *social* and *motivational* dimension that escapes examination in the literature from cognitive psychology. Students in the testing class came to refer to the weekly tests as 'quizzes'. This language modifier, coupled with the descriptions students offered in the interviews, showed

that when freed from the summative assessment culture that can dominate classrooms (Black and Wiliam, 1998: 17-18), students could become intrinsically motivated by assessment. Rather than being regarded as something done 'to' the students, the tests, through regular use, changed the students' behaviour and motivation towards them. In essence, the regular low-stakes tests revealed themselves to be a 'cultural script': a place of socially negotiated behaviour and meaning for the group of students in the testing class and for individuals who negotiated the tests in their own ways as seen in the interview responses.

The change in student motivation also helped to shape behaviour after the initial research period. An assignment was set for both classes to complete in the holiday period and they were told that work would be marked and returned to them via email before their new teacher started teaching in January. Assignments were sent over the holiday period from the students in the testing class only, seeming to suggest a higher level of self-directed motivation than in the non-testing class.

The interviews also revealed a social dimension that had been unexamined in the cognitive psychology literature and the social effect of family life on learning. All students in the samples reported discussing history at home, either at dinner or in general conversation when they found something interesting. The generic question, 'what did you do in history today?' is a form of testing by the parents that had not been considered when planning the project, and it was also clear from the comments that parents were willing to help prepare their children by 'quizzing' them before they had any tests. In some cases, these conversations seem to allow for a mixture of substantive and procedural historical knowledge as indicated by Student 1:

> *I talk a lot about history at home, because I like to talk about monarchs, and why they did what they did, and what happened in certain points, and why the religion was changing, and what people's beliefs were about the religions.*

Explanation, identifying chronology, possible turning points and historical change are reported in this brief answer. The potential of parents to be more than just the recipients of assessment data and act as catalysts for more nuanced historical understanding is currently beyond the scope of the literature on testing mentioned above. Because the literature on the 'testing effect' is based around college students presumably living near, or on campus, in shared accommodation, the need to consider external influences on learning has been minimised. The potential for parental involvement is, however, something that Burnham and Brown (2004) have suggested before. Taking account of the context of the School, this particular external social factor seemed to provide additional learning and testing opportunities that were also low-stakes and helped students to retain information.

The social effects of assessment systems also call into question some of the reported scores in the the second and third assessment for the non-testing class. These pieces, as was the essay before, were essentially summative assessments in that the marks were primarily used for recording, reporting and monitoring purposes. As Harlen and Deakin Crick (2003) outline, this can have an impact on performance. For the non-testing class, working within the usual summative assessment system of the School, the previous scores on the first essay may limit their ability to engage with the task. As Black and Wiliam suggest (1998a), pupils who encounter difficulties and poor results in relation to others, label themselves as failures so they 'quit' in order to avoid further disappointment (1998: 9). Although there were no reports of 'quitting' during the interviews with the non-testing class, the possibility that their

motivation was affected should not be discounted when viewing the performance gap between both classes which was slightly larger than the gap in the first piece.

The essays for the common assessment task were not as developed in the first essays for both classes and this was revealed in the coding process related to substantive knowledge where the disparity between classes was reduced. The reduction of categories or factors in the written discussion seems to suggest that the marks allocated to this piece (ten marks rather than the twenty for their first essay) meant that the students were strategic in their answers. The question in the common assessment demanded more of an argument and this was seen in both classes, indicating the difficulty involved in comparing assessment data across assessment pieces with changes in language. There is a difference greater difference between the results than in the first piece of assessed work but the overall quality of responses was lower which complicates the belief that testing leads to greater retention and deployment of knowledge. The students in the testing class did perform at a higher level than the non-testing class but overall, the quality of what was produced was lower.

The literature review also suggested that even after a delay of months, a significant gap still existed between the 8th Grade US history students who had taken tests and the students who restudied the material or had no review at all. The end of year exam data showed a rather mixed picture. The assessment scores gained by the students in the testing class are *marginally* superior to the non-testing class, challenging the view that testing produces superior learning and performance. This is demonstrated by the brief analysis conducted on the script produced by Student 4 in the testing class. Student 4 was able to recount a great deal of substantive knowledge in the paper that mirrored the first essay, but they were not

able to integrate this with the procedural knowledge needed to address the question and the sources provided. These results seem at odds with Donaghy's (2014) findings regarding the GCSE answers his students produced, as he found that they were able to use their knowledge effectively in conjunction with the sources provided.

The source question on the Tudors seemed to create a levelling effect due to the use of a novel picture. The students in both classes achieved around half marks. The matter of testing securing knowledge is further complicated when considering the final section of the paper. The topic of the Stuarts had recently been taught to the class by the new teacher and the average mark for the testing class on this section was 13.1 compared with 13.9 (out of twenty) for the non-testing class. No definitive answer can be provided for the difference but the question remains as to what led to the change in the final section. With reference to Barnett and Ceci's taxonomy, the 'near' temporal transfer captured by the examination result in the Stuart section reverses the position of the classes. When the 'far' temporal transfer in the Tudor section was not as successful as the 'near' transfer in the common assessment task or the first essay.

Weaknesses

Listening to the interviews in the first instance started to reveal an increasingly complex web of factors involved in the tests and how knowledge was secured. One particular phrase which caused a re-evaluation of the effectiveness of low-stakes tests as the method to secure knowledge and help students deploy it: '*When I learn dates, I think of a football player's number and tray and link [the date] with that'.* No other student had mentioned this technique when describing their general study habits or in relation to their weekly

preparation for the tests. Despite the initial concern to investigate whether testing increases the storage and recall of knowledge, the explanation from the student showed that they were able to secure knowledge via a different route that was *recognised* by the use of the low-stakes test. The mention of the mnemonic device caused a rethink of the project and a return to the literature and even the cognitive psychologists considered mnemonic devices as something of an oddity and that the research on their effectiveness in learning and recalling sequences of knowledge was not widely used (Roediger, 1980; Wang and Thomas, 2013; McCabe, 2015). Recalling other historical periods where memory was prized suggested a deeper examination of their cultures as a possible response to the problem posed by the student's response.

Writing around 55 BC, Cicero in *De Oratore (1942)*, recounts the story of Simonides of Ceos (depicted in a medieval text in Figure 7) who was the inventor of a particularly powerful mnemonic device due to tragic circumstances. Invited to eat at the house of a wealthy nobleman, Simonides was asked to go outside at some point during the meal to meet two young men who had requested his presence. As he looked for them outside, the roof of the hall he had been dining in collapsed, killing all the guests and leaving their bodies destroyed in such a way that they could not be identified. Simonides, however, was able to identify each of the guests because he could recall the specific positions in which they had all been seated, thereby allowing each of them to receive a proper burial. The episode impressed upon Simonides that the best aid to memory consists of orderly arrangement. According to Cicero:

> He inferred that persons desiring to train this faculty must select localities and form mental images of the facts they wish to remember and store these images within these localities, with the result that the arrangement of the localities will preserve the order of the facts. (1942: 467)

This description of the 'method of loci' or the 'memory palace' technique proved useful to scholars in Europe and England through the centuries and helps modern 'mental athletes' in memory competitions perform feats of amazing recall in modern competitions. Despite the longevity and usefulness of mnemonics, they appear difficult to study because they cannot be observed, unlike the action of a test. Carruthers' (2000) work on medieval memorial culture laid bare the problem with the original idea of simply investigating the effects of *testing*. Remembering the things that had been learned was not just an additional task for the medieval scholar, it was something that was deliberately designed and taught alongside the specific content that they were seeking to master. This was done, according to Carruthers, because without a structure, memory is,

> …like an uncatalogued library, a useless contradiction in terms. For human memory should be most like a library of texts, made accessible and useful through various consciously applied heuristic schemes (2000: 39).

Carruthers goes on to suggest that what appears, to modern eyes, as strange medieval marginalia was provided by scribes as devices to help scholars recall what was on the page by linking it to the image provided.

Without constructing some form of heuristic or practical device to help the students store the knowledge that was taught, it left to chance their ability to access that knowledge easily. Essentially, the teaching and testing was helping them to create an uncatalogued library of historical knowledge; something that would not be contemplated when outlining content to be taught either in terms of planning a scheme of work or in the design of a lesson. In this sense, the position was created where the teaching provided was similar to Gramsci's (1971), 'mediocre teacher': someone who could help students become 'informed' by presenting the

'baggage' or 'equipment' of concrete facts but not providing the necessary structure that would enable them to become *educated* (1971: 36).

Conclusions

Education covers a range of practices and ideas and the collective desire to improve and enhance the social, economic and psychological opportunities available to students means that educators often divide the totality of education into discreet spheres. Doing so allows for greater understanding of these discrete sections of social life and the possible identification of actions that not only enhance the knowledge in these discrete sections, but also in the totality. Considered within this context, low-stakes testing is an interesting possible, and partial, solution to the ongoing conversation regarding improvement in education, but it is also a potential problem. The promising data reported in the lab studies are - by their very design – limited, but the drive for results and improvement removes the nuance in order push forward a systemic change agenda in education by the post-2010 government. Cautious statements, once carefully couched within the specific parameters and academic disciplines of the original research, now appear to be repackaged as a total solution to all the educational ills. The transition - from the particular to the universal - is made possible by making science the arbiter of knowledge. The mass of charts and statistical tests within the cognitive psychology literature appear convincing enough, yet we should be more cautious. Feyerabend argued that providing science with the 'sole rights' for dealing in knowledge and removing the contribution of other methods (Feyerabend, 1975: 3) is limiting. This builds on Kuhn's argument (1962) that science is also subject to historical change and social forces so what may be held as 'truth' in the present can change and lead to limiting paths of exploration. Bruner (1990) details this precise problem in his consideration of the earlier

educational 'Cognitive Revolution' of the 1950s. At that point, the search for how humans made sense of the world eventually led to research and actions becoming 'technicalised'; that is to say, a 'stimulus-response' model came to be used in which it was assumed that the brain operated in a similar way to a computer: a specific action would lead *automatically* to the expected outcome (whether that was to the storage of information or to a particular reaction).

That we now accept that the brain and the mind work in more complex ways should remind us of the problem that comes with simple and generalised answers. When considering the findings of the research project, learning and assessment are conditioned by social interactions within, and outside of, the classroom and this should be considered in any explanation. Yet the stress on testing as the means for improving knowledge recall and deployment can lead to the discrete and helpful area of exploration becoming reified into a general and formalised method of education. The picture presented by advocates of low-stakes testing being the best means of enhancing educational improvement is revealed to be more complex when placed within an actual classroom setting. Low-stakes testing *can* help to secure and promote student recall of key information in historical answers that are similar to the original tests and that are taken relatively quickly after the initial teaching and testing. This particular phenomenon deserves further exploration, especially when considering assessment design in Key Stage 3 and the current vogue to show 'progress' in learning. However, in response to the initial concern about the desire to help improve student performance and recall over a longer period of time, the findings and analysis suggest that the use of low-stakes tests *are not sufficient to secure historical knowledge* and effect 'far transfer' of knowledge. As a result, a cautious and carefully considered approach to the use

of cognitive psychology in history classrooms should be the adopted. In addition, it helps to be mindful that examples of successful 'near' temporal transfer should not be read as universal 'far' transfer.

One way to think about the testing effect is to liken it to shooting practice in basketball. In the relative calm of the gym, repeated shots at the basket and the feedback gained will improve the possibility of success over a series of days and weeks. However, in the context of a game that really matters and with emotions running high, the likelihood of your shooting skills matching the gym work is liable to decrease. You can maximise your chances of success if you work on your shooting in different contexts and in practice games, but that requires teamwork, fitness, motivation and coaching. Shooting practice, like regular low-stakes tests, is not bad, but it needs to be accompanied by a whole host of other things in order for the player to perform at a high-level when it counts.

McCrory (2015) demonstrates the need for a more complex understanding of historical knowledge through an inferentialist frame; concepts such as peasant have a short-hand representational meaning *and* a background. The use of the same word or fact by students is not uniform because some carry a rich network of bonds that make the use of the concept more powerful than it is for others (2015: 40-41). In McCrory's work, students may appear to hold relevant knowledge but they *cannot use it* effectively. Her solution is to help build up the rich network of meaning related to substantive knowledge rather than relying *solely* on the 'give and check got' approach to teaching which low-stakes testing has been used for (2015: 42). Low-stakes testing may be helpful when recalling substantive historical knowledge, but

it is not the sole mechanism at play when knowledge is used effectively in 'near' or 'far transfer' contexts.

This point is reinforced by Cain & Chapman (2014) when they argue that extensive substantive background knowledge is necessary if students are to interrogate sources but it is not *sufficient:* completing this task also involves a concept of evidence, modes of reading and interrogation that go beyond factual recall (2014: 117). However, the complexity of such an understanding is not politically palatable or easily communicable when faced with the pressure to improve student performance in a context of accountability and measurement (Biesta, 2010). Adorno (1993) argues that this is a symptom of a system that is not really concerned with education (*Bildung*) but with a 'thin' version or 'half-education' (*Halbbildung*), which focuses solely on assessment grades and simple changes to pedagogy but actually impedes memory and the difficult steps involved in its creation. The findings above show that creating and using historical knowledge is a difficult process which involves a wide range of factors. Ignoring this process for the sake of political expediency diminishes the possibility of being an educated person, the very thing that the advocates of low-stakes testing want to avoid. One-sided explanations of learning are a distortion in the sense that they are 'half-truths' and it is unfortunate that the desire to improve education by using interesting and novel ideas from cognitive psychology tends to remove the complexity inherent in them and to ignore the rich traditions associated with learning within our subject. And that is what makes teaching history great; the openness of the subject and the rich material at hand, should we choose to look for it, helps us to see that working with individuals is a complex business that demands more than a single guaranteed approach.

Implications for practice

The students in the history classes still have problems in being able to draw on their learning months after a topic has been taught, but the dissertation research has demonstrated that to help students enjoy and learn history, multiple approaches are necessary. The plan for next year is to consciously explore and create, in Quintilian's words, a 'treasure-house of eloquence' by deliberately designing the 'storage' of memories and remembering that the human aspect in terms of interactions and learning is just as important. This will be done through the use of 'memory palaces', mnemonics and active learning strategies as well as trying to help students create rich meanings for substantive knowledge over a period of time. The point is not to overwhelm the students with a body of historical substantive knowledge but to help them gain agency and a degree of control through the present. This will take the form of activities that help them to create appropriate storage structures and and use the knowledge they have in a variety of social contexts. Low-stakes testing will continue to be used in order to help students recall key information but it is now understood as one tool in the historical education tool-box.

The dissertation has also brought to the fore the need to be more explicit in teaching procedural knowledge and explaining that history is more than just a collection of facts. To do this, department schemes of work will be adapted to have regular referrals to procedural knowledge and the new approach will be tested with Year 7 this academic year. The plan is to also invite new parents to a sample history lesson next year where they can experience first-hand the lessons their children participate in and to help them, in the discussions at home, to move beyond what they remember of history lessons and to engage in the practise of storing knowledge by being part of the ongoing and present conversation of history.

Collaboration

Elements of the research project have been shared with a variety of educators. Initial discussions about the project within the department occurred in the academic year before the research data was collected. At the end of the research phase, initial findings were shared with colleagues at a department meeting, leading to an extended discussion around more use of regular testing and a greater focus on teaching second order concepts due to a review of the source questions. The feedback provided was positive and concluded with a further discussion on the benefits of a traditional pedagogic and subject-orientated approach in opposition to the School's move towards a whole-school and subject agnostic approach. A training session for all staff at the most recent place of employment drew on the research and had an immediate impact on the School, as the slides were used in tutor sessions, Year group assemblies and teachers in different subjects reviewed their lesson plans to not only think about the use of low-stakes testing to supplement existing teaching tools but to think very carefully about subject-specific constructions of knowledge.

The work was also presented at two events attended by teachers from other schools across the country. The first event was held in May and took place in the South East of England. Over twenty delegates from a variety of state and independent schools attended the session and again the feedback was positive. A Religious Studies teacher from Norfolk had also begun to explore the 'testing effect', and details of the research findings and additional literature were shared in the discussion after the presentation. The final event took place in July at the Schools History Project conference in Leeds. Sixteen delegates signed up for the event representing schools, publishers and Higher Education institutions including University

College London and Manchester Metropolitan University. There were a number of questions around the use of the low-stakes tests and the written work which provoked further clarification and extended discussion with a number of delegates over the duration of the conference. Finally, an article based on aspects of the research and development project has appeared in the September issue of *Teaching History*. The inclusion of this piece will enable the research to be part of the professional discourse of history teachers.

Bibliography

Adorno, T. (1993) Theory of Pseudo-Culture (1959) *Telos*, 1993 (95), 15-38

Aldrich, R. (1990) Introduction, in P. Aldrich (ed) *History in the National Curriculum* London: Kogan Page

Altrichter, H., Posch, P. and Somekh, B. (1993) *Teachers investigate their work* London: Routledge

Bangert-Drowns, R.L., Kulik, C.-C, Kulik, J.A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213-238.

Barnett, S. M. & Ceci, S. J. (2002) 'When and where do we apply what we learn? A taxonomy for far transfer' *Psychological Bulletin*, 128 (4), 612-637

Biesta, G. (2010) *Good Education in an Age of Measurement* London: Paradigm Publishers

Biesta, G. (2012) Mixed Methods in J. Arthur, M. Waring, R.Coe, & Hedges, L (eds.) *Research Methods and Methodologies in Education*, London: Sage

Black, P. & Wiliam, D. (1998a) *Inside the Black Box: Raising standards through classroom assessment*, London: Department of Education and Professional Studies, King's College London

Black, P. & Wiliam, D. (1998b) Assessment and Classroom Learning, *Assessment in Education: Principles, Policy & Practice*, 5 (1), 7-74

Black, P. (2015) Formative assessment – an optimisitic but incomplete vision, *Assessment in Education: Principles, Policy & Practice*, 22 (1), 161-177

Bruner, J. (1990) *Acts of Meaning*, London: Harvard University Press

Burn, K. (2015) The Gove Legacy in the Curriculum: The Case of History in M. Finn (ed.) *The Gove Legacy* London: Palgrave Macmillan

Cain, T. & Chapman, A. (2014) Dysfunctional dichotomies? Deflating bipolar constructions of curriculum and pedagogy through case studies from music and history, *The Curriculum Journal*, 25 (1), 111-129

Carpenter, S. K., Paschler, H. and Cepeda, N.J. (2008) 'Using tests to enhance 8th Grade students' retention of U.S. history facts' *Applied Cognitive Psychology* 23 (6), 760-771

Carruthers, M. (2000) *The Art of Memory*, Cambridge: Cambridge University Press.

Christodolou, D. (2013) 'Research Ed 2013 – How can we discover the root causes of successful teaching and learning?' *The Wing to Heaven*. Available from <https://thewingtoheaven.wordpress.com/2013/09/08/research-ed-2013-how-can-we-discover-the-root-causes-of-successful-teaching-and-learning/>[Accessed 28 July. 2016]

Christodolou, D. (2014) *Seven Myths About Education* Abingdon: Routledge

Cicero (1942 edition) *De Oratore* Translated by E. W. Sutton, London: William Heinemann

Cohen, L. and Manion, L. (1989) *Research Methods in Education*, 3rd ed. London: Routledge

Cox, R. W. (1981) 'Social Force, States and World Orders' *Millennium* 10 (2), 126-155

Cunningham, J. (2014) 'Hegemony' in D. Coughlan & M. Brydon-Miller (eds.) *The SAGE Encyclopaedia of Action Research* SAGE: London

Dawson, I. (2015) '40 Years On: Episode 1' *Thinking History*. Available from <http://www.thinkinghistory.co.uk/diary/2015/01/40-years-on-episode-1/>. [Accessed 28 July 2016]

Department for Education Draft National Curriculum (2013) Available from: <https://www.education.gov.uk/consultations/downloadableDocs/National%20Curriculum%20consultation%20-%20framework%20document%20(2).docx.> [Accessed 28 July 2016]

Didau, D. (2013) 'Testing & assessment – have we been doing the right things for the wrong reasons?' *The Learning Spy*. Available from <http://www.learningspy.co.uk/assessment/afl-have-we-been-doing-the-right-things-for-the-wrong-reasons/>. [Accessed 28 July 2016]

Didau, D. (2015) *What if Everything You Knew About Education was Wrong?* Carmarthen: Crown House Publishing

Donaghy, L. (2014) 'Using regular, low-stakes tests to secure pupils' contextual knowledge in Year 10' *Teaching History, 157, Assessment Edition,* 44-51.

Elliott, J. (1991) *Action Research for Educational Change* Buckingham: Open University Press

Elwood, J. & Murphy, P., (2015). Assessment systems as cultural scripts: a sociocultural theoretical lens on assessment practice and products. *Assessment in Education: Principles, Policy & Practice*, 22 (2), 182–192.

Evans, R (2013) 'Michael Gove's history curriculum is a pub quiz not an education' *New Statesman*, 21 March 2013. Available from <http://www.newstatesman.com/culture/culture/2013/03/michael-gove%E2%80%99s-history-curriculum-pub-quiz-not-education>.[Accessed 28 July 2016]

Feyerabend, P. (1975) *Against Method* London: Verso

Fletcher-Wood, H. (2013) 'What I learned from learning the periodic table and other thoughts on memory and retention of historical knowledge and understanding' *Improving Teaching*. Available from: <https://improvingteaching.co.uk/2013/11/20/what-i-learned-

from-learning-the-periodic-table-and-other-thoughts-on-memory-and-retention-of-historical-knowledge-and-understanding/ >. [Accessed 28 July 2016]

Foer, J. (2011) *Walking with Einstein,* London: Penguin.

Fordham, M. (2015), 'Curriculum Theory, Educational Traditionalism and the Academic Disciplines: Reviving the Liberal Philosophy of Education in Knowledge and the Curriculum' in J. Simons and N. Porter (eds.) *Knowledge and the Curriculum: A collection of essays to accompany E. D. Hirsch's lecture at Policy Exchange.* Available from: <http://www.policyexchange.org.uk/images/publications/knowledge%20and%20the%20curriculum.pdf> [Accessed 28 July 2016]

Fordham, M. (2016), 'Realising and extending Stenhouse's vision of teacher research: the case of English history teachers', *British Educational Research Journal,* 42 (1), 135-150

Friere, P. (1996) *The Pedagogy of the Oppressed* London: Penguin

Gibb, N (2015) 'How E. D. Hirsch Came to Shape UK Government Policy' in J. Simons and N. Porter (eds.) *Knowledge and the Curriculum: A collection of essays to accompany E. D. Hirsch's lecture at Policy Exchange.* Available from: <http://www.policyexchange.org.uk/images/publications/knowledge%20and%20the%20curriculum.pdf> [Accessed 28 July 2016]

Gove, M. (2009) 'What is education for?' Speech to the Royal Society for the Arts. Available from: <https://www.thersa.org/globalassets/pdfs/blogs/gove-speech-to-rsa.pdf> [Accessed 28 July 2016]

Gramsci, A. (1971) *Selections from the Prison Notebooks*, London: Lawrence and Wishart

Giroux, H. (1988) *Teachers as Intellectuals: Toward a Critical Pedagogy of Learning*, London: Bergin and Garvey

Hall, S. (1996) Gramsci's relevance for the study of race and ethnicity, in D. Morely & K-H. Chen (eds.) *Stuart Hall: Critical Dialogues in Cultural Studies,* London: Routledge

Hammersley, M. (2012) Methodological Paradigms in Educational Research. Available from: <https://www.bera.ac.uk/wp-content/uploads/2014/03/Methodological-Paradigms.pdf> [Accessed 28 July 2016]

Hammond, K. (2014) The Knowledge that 'flavours' a claim: towards building and assessing historical knowledge on three scales' *Teaching History, 157, Assessment Edition,* 18-25

Harlen, W. & James, M. (1997) Assessment and Learning: differences and relationships between formative and summative assessment, *Assessment in Education: Principles, Policy & Practice,* 4 (3), 365-379

Harlen, W. & Deakin Crick, R. (2003) Testing and Motivation for Learning *Assessment in Education: Principles, Policy & Practice*, 10 (2), 169-207

Herman, E. S., & Chomsky, N. (1988). *Manufacturing consent: the political economy of the mass media* New York, Pantheon Books

Hirsch, E. D. (1987) *Cultural Literacy: What Every American Needs to Know*, New York: Random House

Hirsch, E. D. (1996) *The Schools We Need: and why we don't have them*, New York: Anchor Books

Hirsch, E. D. (2006) *The Knowledge Deficit: closing the shocking education gap for American children*, Boston: Houghton Mifflin

Hirsch, E. D. (2015) Policy Exchange's Annual Education Lecture. Available from: <http://www.policyexchange.org.uk/modevents/item/the-annual-education-lecture-with-e-d-hirsch> [Accessed 28 July 2016]

Kemmis, S., McTaggart, R. & Nixon, R. (2014) *The Action Research Planner* London: Springer

Kirby, J. (2013) 'What we can learn from Core Knowledge and E.D. Hirsch?' *Pragmatic Education.* Available from <https://pragmaticreform.wordpress.com/2013/01/19/hirsch/> [Accessed 28 July 2016]

Kitson, A., Husbands, C. and Steward, S. (2011) *Teaching and Learning History 11-18* Maidenhead: Open University Press

Kirby, J. (2013) What we can learn from Cognitive Science and Dan Willingham? *Pragmatic Education.* Available from <https://pragmaticreform.wordpress.com/2013/01/26/willingham/> [Accessed 28 July 2016]

Kuhn, T. (1962) *The Structure of Scientific Revolutions* London: University of Chicago Press

Lee, P. and Ashby, R. (200) Progression in Historical Understanding among Students Ages 7-14 in P.Sterns, P. Seixas and S. Wineburg (eds) *Knowing, Teaching and Learning History* New York: NYU Press

Lee, P. J. (2005) Putting Principles into Practice: Understanding History in M. S. Donovan S and J. D. Bransford *How Students Learn* Washington: The National Academies Press

Luff, I. (2016) Cutting the Gordian Knot: taking control of assessment, *Teaching History, Feedback Edition*, 167, 38-45

Lyndon-Cohen, D. (2013) 'A Response to the Proposed National Curriculum in History', 24 February. http://www.historyworkshop.org.uk/nationalcurriculum/

McCabe, J.A. (2015) 'Location, Location, Location! Demonstrating the mnemonic benefit of the method of loci', *Teaching of Psychology,* 42 (2), 169-73.

McCrory, C. (2015) 'The knowledge illusion: who is doing what thinking', in *Teaching History, 161, Support and Independence Edition,* 40-41.

Peal, R. (2013) 'Progressively Worse'. Available from <http://www.civitas.org.uk/publications/progressively-worse/> [Accessed 28 July 2016]

Philpott, J (2010) Assessment, in I. Davies (ed) *Debates in History Teaching* London: Routledge

Quintilian (1920) *Instituto Oratoria* Translated by H E Butler, London: William Heinemann

Radice, A. (2016) 'Memory and Liberal Education' *The Traditional Teacher*. Available from <https://thetraditionalteacher.wordpress.com/2016/06/24/memory-and-liberal-education/> [Accessed 28 July 2016]

Roediger, H. L. (1980) 'The effectiveness of four mnemonics in ordering recall', *Journal of Experimental Psychology: Human Learning & Memory,* 6 (5), 558–67;

Roediger, H. L. and Karpicke, J. D. (1996) Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention *Psychological Science* 17 (3), 249-255.

Saldaña, J. (2009) *The Coding Manual for Qualitative Researchers* London: Sage

Wang, M.H. and Thomas, A. Y. (2013) 'Looking for long-term mnemonic effects on serial recall: the legacy of Simonides', in *American Journal of Psychology* 113 (3), 331-40.

Woodcock, J. (2011) Causal Explanation, in I. Davies (ed) *Debates in History Teaching* London: Routledge

Appendix A – indicative low-stakes test questions

- When did Henry marry Anne Boleyn?
- What does the word ‘annul’ mean?
- Who translated the New Testament into English in 1525?
- What is the process called when the bread and wine become the blood and body of Jesus?
- What is the name given to the idea that the king rules the land and the church?
- When did Henry VIII become king?
- What year did Thomas Cromwell send his investigators to the monasteries?
- Who was the Pope Henry asked to annul his marriage?
- What was the name of the place ‘between’ Heaven and Hell?
- What was the name of the parchment that reduced your time in the place between Heaven and Hell?

Appendix B – Student 1 First Essay

Why did Henry VIII break with Rome? – Student 1

The king of the time of 1533 was Henry VIII, Henry wanted lots of things, a son, power, love, music and many other things. However, to do some of these things, there was something stopping him. For Henry to have a son, he needed a wife, which he had, his wife needed to be able to become pregnant and be strong enough to deliver a healthy and strong child that would be able to pass through birth and stay alive, however the bible stated, ' If a man shall take his brother's wife, it is an unclean thing. They shall be childless.' So, Henry's wife Catherine, was married to Henry's brother before his death, therefore Henry was led to believe, his marriage with Catherine was cursed, so he couldn't have a son. To get a new wife, he would need to divorce with Catherine and find a new wife, which was likely to be his mistress at the time Anne Boleyn, who was pretty, young, capable of being pregnant and promised Henry a son. However the Roman Church, who were the head of divorces, didn't allow Henry to divorce as the Pope Clement VII was a hostage of Catherine's nephew Charles V, who said the divorce was a plot against the family. So, Henry decided, he needed to find a way to get a divorce. The best way to get the divorce, in Henry's ideas, was to break away from the Roman Church. There were 3 reasons he wanted to break away; Love and a Son, New Thinking and Abuses.

The most important reason why Henry broke away from the church was the fact that he wanted love and a son. For example, Henry and Catherine had "fell out of love' and Catherine wasn't giving the son that Henry wanted, so he took action and asked for a divorce. Henry

had also fell in love with a young, attractive woman who was capable of being pregnant again, who had promised him a son, this was Anne Boleyn. With Anne giving Henry a son, Henry could be happier, knowing his family dynasty was safer and more likely to carry on. This is the most important reason to why Henry broke away from Rome because the church in Rome were the only people stopping him from getting his divorce, marrying Anne and having a child that Henry longed for.

The second most important reason why Henry broke away from Rome was New Thinking. If Henry was to break away from the Roman Church he would be able to make up his own, protestant church, which he would be able to control and if he controlled the church, he could get his love life nailed down and sorted out. With a divorce, a marriage and a possible son in his mind, Henry would have certainly wanted to be able to control the church. Another part of New Thinking is Caesaro-Papism, which is where the King can control the land of England, but also the Religious side of England, which links with having his own protestant church. This is the second most important reason to why Henry broke away from Rome because if he did break away, he needed to know what to do and with this as one of his priorities after the break away he would be able to get his Current and future marriages sorted out and he would gain more power over England all together.

The third reason why Henry broke away from the Church in Rome was Abuses. The Roman church was abusing its power in England for the worse. It was selling indulgences, which were pieces of parchment, that sped people time in purgatory up, they were supposedly signed by the Pope , but were easily forged and could be sold by anyone, for a high price. The church was also selling relics, which were objects or body parts that belonged to holy people including Saints and Prophets, the mysterious part of this was that in 1535, Thomas Cromwell found enough parts of the holy cross to make a new one, in one church, while many other churches claimed to have the same cross, so this was obviously a fake relic. This disappointed many of the public and led to rebellions against the church. The church was also doing wrong because the monks weren't obeying the laws that were set in the church, such as law of silence where they wouldn’t tell anyone about what they were doing in the church and they were inviting women into the church and in love with women, which was forbidden in the churches. This is the third reason why Henry wanted to break away from the church because

with this he could get the english public to stand behind him and would make his case stronger due to the fact that he could use these things to go against the pope saying his churches weren't doing religious things and weren’t doing their proper duties all of the time.

Overall, the most important reason for Henry making the break with Roman is his Love for Anne Boleyn and his wanting for the son because he probably wouldn't have needed or wanted a change to the church without these wants, also he was falling out with Catherine and Anne was already his mistress and if he got rid of the church, he would have his way, be able to divorce Catherine and marry Anne.

Appendix C – Mark Scheme for the Common Assessment

1. Use Source A **and** your own knowledge.

Why was the report in Source A sent to the King in 1536?

(5 marks)

Source A

'The Abbot of Fountains has greatly ruined his abbey, wasting his woodland and keeping six mistresses. Six days before our arrival he committed theft…

Extract from a report by the Commissioner sent to visit Fountains Abbey by the King in 1536.

Level	Descriptor
L1 (1-2 marks)	Surface use of source for general assertions/ general statements of context without using the source.
L2 (3 marks)	Asserts general purpose with no development.
L3 (4 marks)	Valid explanation of purpose using either details from the source/ contextual knowledge.
L4 (5 marks)	Valid explanation of purpose using details from the source and contextual knowledge.

2. 'The most important reason that Henry VIII broke from Rome was because he wanted a divorce'.

How far do you agree with this statement?

Use own knowledge to explain your answer. (10 marks)

Level	Descriptor
L1 (1-3 marks)	General answers lacking own knowledge.
L2 (4-6 marks)	Describes reasons using some own knowledge- other reasons may not be mentioned.
L3 (7-8 marks)	Explains several reasons in relation to the questions using own knowledge
L4 (9-10 marks)	Explains several reasons with a conclusion. Top mark will assess the importance of the divorce as a reason in the conclusion. .

Appendix D – Student 1 Response to the Common Assessment

1. Why was the report in Source A sent to the king in 1536?

The report in Source A was sent to the king, Henry VIII in 1536 because it was 1536 where Henry demanded Thomas Cromwell to send commissioners to all of the monasteries in England and Wales. He wanted the commissioners to go to the monasteries because he wanted their land, which was a quarter of England and Wales. However he didn't just want their land, he wanted the income they had, as a monastery's income was greater than that of the king. He had spent so much on wars and displays of splendour, the money he inherited from his father had taken a large dent by 1535. With the monasteries shut down he would be able to take all of their income after setting up his own church. Henry wanted Cromwell to find evidence that the monasteries should be shut down, to give to parliament so he would get them to support him in his want to shut down the monasteries, but also help to not cause rebellions over England and war with Spain and Charles V.

The source also shows that the Abbot of Fountains was wasting his woodland, keeping six mistresses and six days before the report was written he committed theft. With this in the source it puts the abbot and the monastery in a bad situation. They were put in a bad situation because the abbots should not be having mistresses. They shouldn't be having mistresses visit them because it is distracting them and maybe even breaking the law of silence in the monasteries, they should use their land properly and to a good measure, not wasting it because they have been given it to help them worship God not to waste it and not even use.

It is also certainly bad that a person with such high social status that works in such a highly looked up to place is stealing. He shouldn't be doing this because it is against God and his commandments and the monastic rules and is a bad role model to others who look up to his job and shameful to the monasteries in general. If the report had been written at a different date then they may not have found as many bad things to do with the monastery. The report was sent to the king to help him control the monasteries, gain wealth, land and power.

'The most important reason that Henry VIII broke away from Rome was because he wanted a divorce' How far do you agree with this statement?

I completely agree with this statement. I agree with this because Henry wanted a son, to carry on his family dynasty. However he was led to believe Catherine couldn't have a child because their marriage was cursed due to Catherine being Henry's brother's Widow and in the Bible it says If a man shall take his brother's wife, it is an unclean thing. They shall be childless. To have a Baby he would need a new wife. If he had a new wife he would be more likely to have the son he wanted because their marriage wasn't cursed. Therefore, Henry would need to divorce Catherine and find a new wife, which was likely to be his mistress at the time Anne Boleyn, who was pretty, young, capable of being pregnant and promised Henry a son.
Before Henry asked, in person, he ordered his Lord Chancellor, Thomas Wolsey, a high ranking person in the Catholic church, to help him. He wanted Wolsey to ask the new pope, Clement VII, to annul the royal marriage. However, this was a tough challenge because the previous pope had said no to the divorce and the popes weren't known to change minds, even across two popes. With that in mind, Wolsey still did what Henry asked, but Clement sent his Cardinal Campeggio to England, where he was to listen to both sides of the divorce and make a recommendation to the pope. After hearing the two people speak, Campeggio left for Rome, without telling them his decision and never came back. So Henry decided to ask again, in person. However the Roman Church, the place where Henry would have to ask for a divorce, didn't allow Henry to divorce Catherine because Pope Clement VII was a prisoner of Charles V, Catherine's nephew. The best way to get the divorce Henry wanted, he needed to break away from the church in Rome.

I feel that Divorce was the main reason Henry broke away from Rome because he wanted to carry on in the family dynasty and he could find new love in Anne Boleyn, as he thought the marriage between Catherine and him had "fell out of love," as Catherine didn't give the son Henry wanted.

Another reason for Henry to break away from Rome would be some new ideas that Henry had thought of, or had been put in his head. An idea that someone had told him about and spoken a lot about was Caesaro-Papism, where the king would rule over the land, but also control the religious side to the country. With Caesaro-Papism, Henry would create his own protestant church, where he would be the person that decides on his divorce and he would be able to divorce Catherine. With the protestant church, Henry would give the people full communication with God and allow them to connect with him in their mother tongue, mostly in English, so the Bible and mass would be in English. People wanted this because they felt the church was blocking their full connection with God and they didn't know how to pray to God and what was going on in the Bible.

Henry also wanted to break away from Rome because the church was abusing his people for money. The Church was abusing the people as they had taxes that could be up to 1/10 of someone's earnings. This was abusing their power because none of the public knew where the church was putting the money and if someone wanted to speak out, against the church they could be punished. The church was also abusing their power by selling Relics, which were holy body parts, belongings or objects which were owned, used or part of a holy person. In 1536, Henry ordered his 'Vicar General' Thomas Cromwell to send some commissioners to monasteries and look for some of the bad parts of the church, in St Edmund's convent, one of the commissioners said " among the relics we found coals that St Lawrence was burnt upon, clippings of St Edmunds toenails, St Thomas of Canterbury's penknife and his boots and enough of the holy cross to build a whole one." This stands out because the church had enough of the holy cross to build a new one, but many other churches around the country also claimed to have some of the cross too. The church was also selling indulgences, which were bits of parchment that were said to help you through purgatory and decrease your time. The church was saying that if you used, owned or prayed with these indulgences, your time

in purgatory would lower and it would be easier for you. The indulgences could easily be forged or fake and signed be a random person, but still be sold for a high price. With these abuses, Henry wanted to get rid of the church because the people didn't like what they were doing, some of the churches were bringing in mistresses, badly using woodland and gambling, which were all against the monastic rules

There are many other reasons why Henry wanted to break away from Rome, but the main reason, in my opinion, was to get a divorce so, he could annul his marriage with Catherine of Aragon, marry Anne Boleyn, then he would be able to have a Baby in a marriage that wasn't cursed so his family could carry on, in their dynasty. Abuses and New thinking were also important but New thinking said with Henry's own church he would be able to get a divorce, which is relating to the divorce. Abuses is different, but it didn't give him a big enough reason to completely break away from the church.

Appendix E – Interview Questions

List of indicative questions to be asked in the student interviews

Initial round of interviews at the start of term

Why do you think History is one of the subjects you study at school?
What kind of things do you need to know in History lessons?
What can you remember about what you studied last year?
Is it worth remembering what you studied? If so, why?
What does the term knowledge mean to you?
How important is factual knowledge in History?
What other kinds of knowledge do you need in History?
What is the best way to learn something?
How do you know when you have learned something?
How can you show that you have learned something?
How do you learn dates and names in history?
How do you learn the sequence of events in history?
Do you read any books on History?
Do you talk about what you have done in History lessons with other people outside of the school?
Do you use tests outside of class?
How do you revise for an exam?
Are tests useful to students?

Why do you think teachers ask you to take tests?

How do you prepare for tests?

Second round of interviews at the end of the term

Essentially the same questions will be asked as in the first round of interviews but a number of additional questions will be asked to those students who took part in the series of informal low-stakes tests during the course of the term:

How useful did you find the tests?

Were you worried before you took them?

Did you enjoy them?

Did you think you got better at them?

Did the tests matter?

Did the tests have an effect on your work in History?

If your results improved over time, why do you think this happened?

If your results did not improve over time, why do you think that was?

Interview with Student 1 in the testing class (second phase).

Appendix F – Interview with Student 1 (second phase)

NRD: So, why do you think history is one of the subjects you study at school?

Student 1: I think we study it because we're looking at the past, of the happenings in England and other countries in the world, and it's very useful to know.

NRD: Okay, and what kinds of things do you need to know in history lessons?

Student 1: We need to know dates, times, people's names, and important facts about people to help us with the history.

NRD: Okay, and what can you remember about what you studied last year?

Student 1: Last year, we studied about William the Conqueror, and the motte-and-bailey castles, and why they built them, to conquer the land.

NRD: Okay, is it worth remembering what you studied last year?

Student 1: Yes, because when we're going to somewhere like xxxxxx Castle, we can realise that that's a motte-and-bailey castle. It's useful to know that that is a motte-and-bailey castle, and when it was built.

NRD: What does the term 'knowledge' mean to you?

Student 1: Knowledge means that you have a certain understanding of what's going on.

NRD: How important is factual knowledge in history?

Student 1: Factual knowledge in history is very important, as it's one of the key parts in history, and if you don't know your facts, you don't really know history.

NRD: What other kinds of knowledge do you know in history or do you need to know in history?

Student 1: Some other facts that you need to know, knowledge that you need to know are maybe dates, and times, and knowing when things have happened, and why they've happened, maybe.

NRD: Okay. What is the best way to learn something?

Student 1: The best way to learn something is to cover it, but then not just cover it once. Go back over it and maybe test yourself every couple of days or weeks, just to check that you know it, and then you can move on.

NRD: Okay, and how do you know when you've learned something?

Student 1: I know when I've learned something when I'm 100% confident with it, and I think I can go into a test and not worry about it, and know that I'm fine.

NRD: How can you show that you have learned something?

Student 1: I can show that I've learned something by, in a test, just making sure that everything's written down, and everything's right. If I get 100%, I've showed that I know everything.

NRD: How do you learn dates and names in history?

Student 1: In history, when I'm learning dates and names, I usually tend to, like, write them down, and think of a key way to remember the date along with the name. So, with the Magna Carta, I can think of King John and his reign, and I know that his reign was somewhere around 1215, and the Magna Carta was in 1215.

NRD: Okay, so just to, kind of, expand here, you're basically saying you make links between..?

Student 1: Yes.

NRD: Okay. How do you learn the sequence of events in history?

Student 1: Sometimes I have, like, a rhyme, or a link to something, because it wouldn't make sense to say, "It was Henry VI before Henry V."

NRD: Do you read any books on history?

Student 1: Sometimes. I've read books about, like, the history of the castles, and places in London and around, like bridges.

NRD: Okay, and do you talk about what you've done in history lessons with anyone outside of school?

Student 1: Sometimes. Sometimes I'm really interested in what we've studied at home, and go home and tell my parents, and discuss about it with them.

NRD: Do you use tests outside of class?

Student 1: Not much, unless we have a big test coming up and I want to study. A big way to help is to test your knowledge after you've maybe revised for a bit.

NRD: Okay, and so how do you revise for an exam?

Student 1: I revise for an exam with flash cards, and obviously going back, testing myself, and if I get it wrong, then I'd study that using the flash cards, or bullet points. Usually, when I'm revising, I don't do it for very long, but so I do it in short periods. So I have, like, 20 minutes, and then a break for, like, 5 minutes. Then, I do another 20, and I usually do that for around three hours.

NRD: Okay. So, just talk to me about the flash cards. So, how do you use them?

Student 1: I would have a bullet point on the flash cards, so I'd have a title, and then I have bullet points with different knowledge. Then, so I'd have the topic, and then I'd get my brother to test me. He would say, "Talk about the topic," and I would try to say what I've got written down on those bullet points.

NRD: Okay. Are tests useful to students?

Student 1: I think sometimes they are, but sometimes people get in a bit of a panic and don't do as well as if they were just doing normal class work. If they didn't understand it was a test, and it was just some normal classwork, they wouldn't feel as pressured. It would be a bit more relaxed.

NRD: Why do you think teachers ask you to take tests?

Student 1: So they can test your knowledge and see that you understand the topic, and then they can move on to further topics.

NRD: Okay, and how do you prepare for tests?

Student 1: Sometimes, I try to calm myself down before I go into the tests, and I just think back to the flash cards, and the tests that I've done with other people. Try to remember as much as that from the flash cards in the tests, and take it into the test.

Interview 2

NRD: Why do we learn history?

Student 1: We learn history so we can understand what's happened in the past, and we can learn culture, and learn more, like how we're here, in some ways.

NRD: What kinds of things do you need to know in history lessons?

Student 1: We need to know dates, people's names that are important, places like Hampton Court, and monarchs really.

NRD: What can you remember about what we studied this term?

Student 1: We studied the dissolution of monasteries, which was where they got rid of all the monasteries and made Henry the head of the church. We learned about Edward's reign. We learned about Elizabeth's reign and Mary's reign, and in those three reigns there are lots of changes to the religion. In Elizabeth's reign, we practiced about what she had to decide on religions and marriage, and if she should get married, and what religion she should take.

NRD: Okay, thank you. Is it worth remembering what you studied this term?

Student 1: It is, yes.

NRD: It is? Why.

Student 1: It's probably worth remembering a lot of things, so just further in life, you can have that knowledge and you can understand what's happened. If someone asks you a question about the subject, then you can know it.

NRD: Okay. What does the term 'knowledge' mean to you?

Student 1: Knowledge, to me, means to understand what's going on, and to have a sort of understanding of what's going on around you, and knowing exactly what's happening in a question. If you're asked a question like, "What was Elizabeth's religion?" The knowledge would be that she was Cath…, no, Protestant.

NRD: Okay. What does the term 'knowledge' mean to you? Sorry, did I just ask that question? Sorry, I'm just losing my place here. How important is factual knowledge in history then?

Student 1: Factual knowledge is very important, so you can understand exactly a date, or a person's name and what number king they were, or queen. It's very important in history specifically because history's mostly about facts.

NRD: Okay. What other kinds of knowledge do you need in history?

Student 1: In history, you need knowledge about numerics, so you can know exactly when the date was, or maybe a birthday, or how long they reigned for. Numbers are quite important in history. There's also knowledge of just general people, and time zones, and…

NRD: No, that's fine. What is the best way to learn something?

Student 1: The best way to learn something is to study it for quite a long time, so you can really nail it down instead of just spending a few lessons on it. When you learn it for longer, then you could probably have a better chance of remembering it, because you've got more time to take on that subject.

NRD: Okay. How do you know when you've learned something?

Student 1: I know when I've learned something if, say, there's a test and that question comes up, and if I can I answer it to a good standard, then I know that I've learned to compared to not knowing anything about the answer.

NRD: Okay. How can you show that you've learned something?

Student 1: I can show that I've learned something by maybe testing myself, or getting someone else to ask some questions about the subject, and then I can show that I've learned it, and…yes.

NRD: Okay. How do you learn dates and names in history?

Student 1: With dates and names, I write them down and try to rhyme them, or try to remember them at least, off by heart, so then I can just nail a name to a date.

NRD: How do you learn the sequence of events in history?

Student 1: To learn the sequence of events is quite hard, but first, I'd say the dates are really important, and then you can maybe learn the first date, and then the first happening, and then just do the second date and second happening, and then try and match them up, and then just do that until you get to the end.

NRD: Do you read any books on history?

Student 1: Not really.

NRD: Not really. No, that's fine. Do you talk about what you've done in history lessons with other people outside of the school?

Student 1: Yes. I talk a lot about history at home, because I like to talk about monarchs, and why they did what they did, and what happened in certain points, and why the religion was changing, and what people's beliefs were about the religions.

NRD: Okay. How do you revise for an exam?

Student 1: When I'm revising for an exam, I like to get pieces of paper, and then write down and try to remember them, and try to understand as much as I can. I read through books, and I spend quite a lot of time, but I don't do it all in one bit, I do it in short breaks. I do 20 minutes and then a break, and carry on like that for a few hours.

NRD: Okay. Are tests useful to students?

Student 1: Yes, because you can understand what you've learned and what you haven't learned. Instead of looking at your score and thinking, "I've got 80%, that's really good," you should look at the 20% and say, "I need to learn on that." Whereas, if you didn't do the test, you wouldn't know that you got that 20% wrong.

NRD: Okay. Are tests useful to – sorry, I just asked that one. Why do you think teachers ask you to take tests?

Student 1: I think teachers ask us to take tests so we can look back on the subject. They know that everyone understands it, and everyone has that knowledge, so you can move on to the next subject.

NRD: Okay. How do you prepare for tests? I know you may have answered this before, but…

Student 1: Yes. I just try to remember as much as possible, and I do it in short breaks. Sometimes, I struggle to do it in the short breaks because I just want to get it out of the way, but I know that's not really the best idea.

NRD: Okay. How useful did you find the test/quizzes that we've been doing?

Student 1: They've been really useful, because if I got something wrong, I can look at the answer. Then, if that question came up again next week, then I know I should be able to get it.

NRD: Okay. Were you worried before you took them?

Student 1: No.

NRD: Okay. Did you enjoy them?

Student 1: Yes, because some of the questions were quite hard, and it made me think about things that I hadn't really thought about before.

NRD: Okay. Do you think you got better at them?

Student 1: Yes, but sometimes as it got on, they got a bit harder. I still found it a bit easier as time went on as well though, because I got into the rhythm of doing a test each Tuesday. Sometimes questions would come up more than once, so they'd come up in the next week. Then, if I know I got that one, the last week's, right, then I know I get it right again.

NRD: Okay. Did the tests matter?

Student 1: Yes.

NRD: Explain. In what sense?

Student 1: Well, they mattered to me because I knew exactly what was going on, and I knew what I'd done and hadn't done right. Then, in a way, they didn't really matter because they wouldn't go towards a grade, or

they wouldn't really threaten us. They wouldn't make us move sets or anything. It's just to get some ground and know what we're doing.

NRD: Did the tests have an effect on your work in history?

Student 1: Well, they affected it in a positive way because if I knew I got something wrong, then I knew that I'd got it wrong, and I knew that I knew the answer from now on. They also affected my work because I got a better understanding of tests, and what happened in questions. Yes, it affected it quite well.

NRD: Okay. If your results improved over time, why do you think this happened?

Student 1: I think this happened because, as I said before, lots of questions got repeated, so people could have an understanding, and really nail it down and know what's going on. I think if you get it wrong one time, and then you get the result back, you can see that wrong answer, and then you've got the answer, and then you know that you've improved on the next time, if you've got it right.

NRD: Okay. If your results did not improve over time, why do you think this might happen?

Student 1: I think this might happen because, well, it's quite hard to nail some things down, but if you didn't get the answers back and you were still clueless on this subject, then you probably wouldn't know what to revise for. You wouldn't really have an understanding of everything. If a question came up, and then you didn't understand it because you hadn't revised for it, then that probably wouldn't improve your…

NRD: Okay, brilliant. Thanks very much Student 1. Let's stop this here.