A Cognitive Approach to Assessment for Learning: Assessing the impact of changing formative assessment practices in a secondary school science faculty

Jessica Tratt

A Research & Development Project
Submitted for the
MSc Learning & Teaching 2018

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	Tratt
First Name	Jessica
Faculty Board	Education
Title of Dissertation	A cognitive approach to assessment for learning: assessing the impact of changing formative assessment practices in a secondary school science faculty.

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.

A Cognitive Approach to Assessment for Learning

Assessing the impact of changing formative assessment practices in a secondary school science faculty

Abstract

This research and development project took place across a science faculty in a UK state secondary school. Following trials, an assessment intervention was introduced to all science classes across key stages three and four in one academic year. The intervention sought to improve on an aspect of formative assessment, regular short tests, by combining literature on formative and summative assessment design and implementation with evidence from cognitive science relating to the nature of learning over time and how to improve it. The design of the intervention came from applying ideas from research to the specific context of the school, science faculty, staff and students at School X.

Quantitative analysis suggests an improvement in examination performance for students of all abilities in years 9 and 10 but a slight decrease in examination performance for students in years 7 and 8. Student and teacher opinion about the intervention is that it benefits learning. The primary reasons for this benefit are identified as improved diagnostic feedback as a result of increasing the inferential capacity of the assessment and improvements in the self-regulatory behaviour of students.

Contents

1. Introduction and Rationale 4

2. Literature Review

2.1 The current state and critiques of formative assessment in schools..... 8
2.2 Linking assessment and learning 20
2.3 Using cognitive science to understand how learning takes place 23
2.4 Improving the validity of teacher and student judgments of learning..... 27
2.5 Self-regulation 31

3. Methodology

3.1 Research questions and approach..... 33
3.2 The intervention: Progress Checks..... 34
3.3 Data collection methods..... 38
3.4 Ethical considerations..... 45
3.5 Collaboration..... 47

4. Findings and Discussion

4.1 Student surveys 49
4.2 Student interviews 57
4.3 Teacher interviews 60
4.4 Summative data 64
4.5 Discussion 68

5. Conclusions and Implications

5.1 Conclusions 78
5.2 Implications for future practice and research 82

6. Reference List 84

7. Appendices

Appendix A: Letter of permission to headteacher 92
Appendix B: Headteacher's signed consent 94
Appendix C: Example of student survey with coding 95
Appendix D: Example section of data analysis from student surveys 97
Appendix E: Student and teacher interview questions 98
Appendix F: Example sections of interview transcripts with analysis 99
Appendix G: Year 7-10 target data 101

Chapter 1

Introduction and Rationale

Assessment and Learning

The relationship between learning and assessment is fraught with difficulty. In education systems worldwide a 'target-driven accountability culture' (Stobart, 2017, p. 4) has risen and, in many countries, prevailed (Baird et al, 2016; Koretz, 2008; Wilson, 2018;). Consequently, external high-stakes assessment becomes highly influential in decision-making in schools and classrooms (Harlen, 2004). Accountability measures that focus on results leads to schools and teachers placing great focus on preparation for exams (Shepard, 2008).

Assessment systems may have always been high-stakes, but the stakes are no longer only high for the individual sitting the exam; they are often equally high for those that do not sit the exams but are held responsible for the results (Barton, 2018). This changes the relationship that headteachers, classroom teachers and students have with assessment and many would argue the educational landscape in the UK has become awash with an obsession for sub-levels, grades, targets and individual tracking (Christodoulou, 2016), all with a focus on what outcomes a student, or cohort of students will achieve in a series of exams at the end of a period of schooling.

This warped relationship changes assessment in classrooms carried out by teachers and, whilst there is an emphasis on using assessment to benefit learning and inform it on a daily basis, the much-hoped for benefits of processes such as Assessment for Learning may have been diminished by the pressure imposed by an unforgiving culture (Bennett, 2011; Crooks, 2011; Swaffield, 2009). It is possible that building stronger links between assessment and

learning, by finding out more about the ways in which assessment contributes to, and aids learning may steer a path that allows teachers to navigate this difficult relationship (Baird et al., 2017; Shepard, Penuel and Pellegrino, 2018a; Stobart, 2008; Wiliam, 2018).

School context

It is not surprising that, over the past seven years that I have worked in schools, it has always been customary to regularly assess students with assessment either labelled as formative or summative. This assessment has usually taken the form of either an end-of-topic test using practice summative material in a 'mock' style paper, or less frequently, extended tasks used to determine a national curriculum level based on a set of criteria (for example, Badger tasks in secondary science). In School X, these tests were spoken of as 'diagnostic' by the faculty and in the wider school, but my experience was that they were flawed in their ability to infer success and failure in learning across a topic. My own observations upon becoming Head of Faculty showed there to be a lack of follow-up and use by teachers in order to respond and improve instruction based on the 'diagnosis'. At times it became a system of simply inputting numbers which were prone to large margins of error that were then very rarely tracked systematically. There was a realisation within the faculty that we were using summative materials in a summative manner and calling it formative assessment. The use of these tests became further challenged by the advent of a new GCSE curriculum in 2016 where we found our previously rich resource of practice summative material was depleted initially to one sample paper from the exam board.

With a relatively inexperienced team who required support in assessment, a school-wide focus on improving diagnostic and formative assessment and an opportunity posed by a changing curriculum, I decided to design new assessment materials with the sole purpose of

having high diagnostic functionality with a wide enough scope to allow teachers to form valid and accurate judgements about the progress of their students. We needed to re-establish strong formative assessment to benefit the progress of students in the faculty but were also aware of the notorious difficulty in the effective implementation of formative assessment, particularly over time. We decided to combine the effort with research from the fields of cognitive science and psychology in order to make the process as effective as possible. The decision to use research from cognitive science to improve education is currently widely discussed across the education community in the UK (Blakemore, 2018) and recently there have been calls from leading academics in the field of assessment to consider how a view of learning from this perspective may benefit and enhance assessment, and conversely, how assessment may be used to improve learning (Wiliam, 2018: Shepard, Penuel and Pellegrino, 2018, Baird et al., 2017).

School X is a non-selective community school situated in South London. The student body has a wide range of backgrounds and abilities. The Science Faculty has eight full-time and three part-time teachers. Many of the teachers (five out of eight full-time teachers) are in their first three years of teaching.

Definitions

Formative assessment/assessment for learning

I use these two terms interchangeably throughout the project. This is a deliberate choice because, although the two remain distinct in some aspects of the literature, my experience is that they are used to mean the same thing in school communities. I use the following definition for both of these terms:

> '*Practice in a classroom is formative to the extent that the evidence about student achievement is elicited, interpreted, and used by teachers, learners and their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited.'* (Black and William, 2009, p. 9)

Cognitive

The use of cognitive in the title refers to all aspects of learning science; including cognitive science, educational psychology, psychology more broadly and neuroscience. There is a potential confusion between using this term as explained and for it to be interpreted as meaning a solely cognitive constructivist approach to learning theory. Whilst many aspects of the project do draw on the work of theorists that can be classified within this orientation, I do not use it to exclusively refer to this, and at times draw on what would be considered behaviourist or social constructivist perspectives alongside research from cognitive science.

Chapter 2

Literature Review

2.1 The current state and critiques of formative assessment in schools

2.1.1 The widespread use of formative assessment in UK schools

Formative assessment, or assessment for learning, is ubiquitous in English schools. As Robert Coe notes, 'It is now a rare thing, in my experience, to meet any teacher in any school in England who would not claim to be doing Assessment for Learning' (Coe, 2013, p.10). In schools today, the manifestations and practices of these ideas stem largely from, and were popularised by, the work of Dylan Wiliam and Paul Black (1998) (Baird et. al, 2014, Sadler, 1998, Black et al. 2004). Their work evaluated the evidence for the positive impact of formative assessment practices and this, along with their consequent work (Black et al., 2003), shaped the trajectory of its implementation in schools in the UK and internationally (Birenbaum et al., 2015; Christodoulou, 2016). Following the publication, supported by the Assessment Reform Group and with the consequent backing of initiatives by government (Department for Children, School and Families, 2008) large numbers of articles have documented this widespread use (Baird et. al, 2014.). These mostly comprise of case studies, theoretical articles and action research rather than quantitative, empirical studies that attempt to replicate the reported measurable effect on learning that Black and Wiliam (1998) described. From this established, and extensive research base it is perhaps not surprising that the ways in which teachers implement formative assessment in their classrooms can vary substantially and with varying quality (Black, 2015, Birenbaum et al. 2015). For example, a recent project aiming to gather teacher views on how formative assessment is implemented at School X (Author, 2017) found that some teachers place greater significance on their own

action and responses to assessment opportunities in classroom, whereas others place more importance on the use of the processes by students. Alongside these potentially different viewpoints from which to implement formative assessment, Swaffield (2009, p. 4) recognises the risk that comes with any large-scale government supported initiatives, where differences can be identified between teachers who are aware of the underlying theoretical principles and those for whom it presents as a top-down, tick-box exercise.

2.1.2 Criticisms of implementation

The hoped-for benefits of implementing formative assessment strategies have not yet been realised (Christodolou, 2016, p. 20). There is little evidence to suggest improvement in attainment as a result of the strategies (Coe, 2013) and further evidence to suggest that government, schools and teachers have not successfully translated theory into practice so that the benefits may be realised (Black 2015; Swaffield, 2009). Wiliam himself has said that 'there are very few schools where all the principles of AfL, as I understand them, are being implemented effectively' (Stewart, 2012) and he has since claimed that it was a mistake to name the process formative assessment 'rather than something like "responsive teaching"' (Wiliam, 2013). Others have contributed widely-cited and -read critical reviews (Bennett 2011; Swaffield, 2009). Bennett, for example, identifies six issues that he argues have undermined the success of formative assessment in classrooms:

1. The definitional issue
2. The effectiveness issue
3. The domain dependency issue
4. The measurement issue
5. The professional development issue
6. The system issue

(Bennett, 2011, p. 5)

The definitional issue can be found in the use of the term formative assessment to mean either process or instrument. The issue is further complicated by the use of different terms that some consider holding different meanings, for example assessment for learning versus formative assessment (Bennett, 2011, p. 7). Bennett takes issue with the polarisation of process versus instrument and notes that *'*even the most carefully constructed, scientifically supported instrument is unlikely to be effective instructionally if the process surrounding its use is flawed' (p. 7), and argues that the same is true for a process where the instrumentation is 'not well-suited for the intended purpose'. Bennett argues that, to be effective, formative assessment should rather be seen as a 'thoughtful integration of process and purposefully designed methodology or instrumentation' (2011, p. 7).

In the effectiveness issue Bennett identifies flaws in the claims that empirical evidence strongly supports the assertion that formative assessment practices greatly improve learning. This argument is reinforced by Stobart and Hopfenbeck (Baird et al., 2014). The authors explain that there is criticism of Black and Wiliam's work because it is often perceived as a meta-analysis, which it is not (p. 45), and cite the work of Kingston and Nash (2011), explaining that their meta-analysis that attempted to measure the effectiveness of formative assessment and assessment for learning 'found a weight mean effect size of 0.28 – well below the often-cited 0.4-0.7 from Black and Wiliam's estimates' (Baird et al., 2014, p. 45). A very recent Education Endowment Foundation publication (2018) provides new evidence, it finds that the Embedding Formative Assessment project led by Dylan Wiliam across 140 UK schools resulted in students making 'the equivalent of two months' additional progress' in their GCSE exams (Education Endowment Foundation, 2018, p. 1). Whilst this provides further evidence for the efficacy of formative assessment strategies, the primary focus of this specific project was the professional development of staff rather than on the strategies themselves.

Regarding the domain dependency issue, Bennett argues that a successful approach to formative assessment must mean a different approach for curriculum areas that deal with different domains of knowledge. He calls for greater attention to cognitive-domain understanding in teacher practice, alongside a more sensitive approach to the 'intellectual tools and instrumentation' used within varying domains (Bennet, 2011, p.15).

The measurement and professional development issues are concerned with the ability of teachers to make useful and accurate inferences from the evidence they gather about student learning in their classrooms. Any assessment process or task in education will always pose a degree of inaccuracy as 'we cannot know with certainty what understanding exists inside a student's head' (Pellegrino et al., 2001, p. 42). Even correctly identifying a lack of knowledge within an individual or group does not necessarily mean that a teacher is able to correctly hypothesise about the cause. Many causes may lead to observable error and it is the teacher's role to provide opportunities that may allow a student to elaborate or identify for themselves the foundations for their misunderstanding. The ability of a teacher to create a culture in which this diagnosis can be made is crucial, and expertise within the cognitive domain is a key component of this ability (Bennett, 2011, p. 17). Bennett also notes that formative assessment does indeed involve assessment and that the measurement of this assessment must be as accurate as possible, recommending that the principles of measurement science be considered alongside developing teacher's expertise in inference.

It is within the educational systems that provide the setting for formative assessment to be used in schools that Bennett's final criticism lies. High stakes accountability testing of students impacts upon assessment practices within classrooms and a coherence between the two

must be built 'from the ground up' if it is to exist effectively (Bennett, 2011, p. 20). Swaffield provides an example of this in the UK, where the Assessment for Learning Strategy employed by government is criticised for primarily focusing on testing and summative assessment with a lack of depth in explanation of the core principles of formative assessment. Swaffield states that 'the AfL Strategy practically ignores the process and pedagogical values and the underpinning principles' (Swaffield, 2009, p. 13).

Despite their criticisms, Bennett and Swaffield are both positive about the potential that formative assessment holds. Bennett describes the principles as having 'considerable promise' (Bennett, 2011, p. 21) and Swaffield (2009) claims that the potential 'prizes are too precious to be passed over lightly' (2009, p. 14). These criticisms of how formative assessment has been implemented not only give reason to not give up on implementing formative assessment in schools but also sets a direction for which it may be improved.

2.1.3 The process of formative assessment

The definitional, domain dependency and measurement issues that Bennett raises are particularly important to consider when translating the theories of formative assessment into practice within a classroom.

The criticisms of formative assessment lead to the conclusion that there is a need for formative assessment in schools to develop into a more coherent process that both teachers and students understand. This can be borne out through the use of well-designed tools for the required domain at the outset of the process, whilst also providing opportunity for

meaningful measurement and response by both teacher and student to the evidence of learning raised by these tools.

The identifiable features of formative assessment adopted in many schools followed those described by Black et al. (2003). The authors describe four key activities that could be used to implement the principles of formative assessment, three of which are important to consider when designing medium-cycle (Wiliam, 2006) formative assessment tasks:

- Classroom questioning (or dialogue)
- Feedback through marking
- Peer- and self-assessment
- Formative use of summative tests

(Black et al., 2003, pp. 69-70)

The feedback through marking theme resulted in teachers beginning to focus on comment marking rather than providing grades. The idea of curriculum focused targets in marking is now well embedded in UK secondary schools but the way in which this is implemented is a point of controversy, particularly in relation to teacher workload (Elliot et. al, 2016). A practice that this original research influenced was a trend for so-called 'triple marking' where teachers and students engage in extended dialogue (often in different coloured pens) through written comments in students' exercise books (Elliot et al., 2016). Since reaching a peak, awareness of marking workload has been raised from the government and Ofsted (Independent Teacher Workload Review Group, 2016) with the latter now announcing no expectations to see any particular length or frequency of written comments in students' books.

The efficacy of providing feedback through written marking specifically has been questioned by Elliot et al.'s review into written marking (2016) which questions the discrepancy between the 'enormous amount of effort invested in marking books, and the very small number of

robust studies that have been completed to date' (Elliot et al., 2016, p. 4). However, the review does support the principle from formative assessment that providing grade-only feedback does not enhance student progress and concludes that the majority of studies (although not all) suggest that grades 'obscure comments' (p. 10) and thus make them less impactful. The review also comments upon the requirement for effective feedback to be highly specific; drawing on a wider evidence base including research into verbal and peer feedback as well as from psychology. The reviewers 'consistently find that the specificity of feedback is a key determinant of its impact on performance' (p. 20), but warn that this increase in specificity makes marking more time-consuming. It references code strategies that have been used by schools to reduce the time spent on writing comments and highlights research that suggests there is 'no difference between the effectiveness of coded or uncoded feedback' (p. 20).

The third of the four cornerstones proposed by Black et al. (2003) is the use of self- and peer assessment and is based on the principal that, unless students know what it is they are aiming to achieve, they will not be easily able to attain the intended learning outcome. The recommendations that came from this strand of the research were threefold: the need for learning outcomes and aims to be very clear to students, with the use of concrete examples where necessary; the importance in teaching students how to self- and peer-assess; and teaching students to approach work with the goal in mind so that they are able to assess their own progress as they proceed through a task. It is the second of these findings that is most pertinent to a project such as the one discussed in this essay and there is an opportunity to realise the recommendations in relation to marking workload if students themselves can become better assessors.

Explaining the fourth strand of formative assessment, Black et al. differentiate between a test with a summative function and a test with a formative function by explaining that summative assessments are usually 'infrequent' and are 'isolated from normal teaching and learning' (Black et al., 2003, p.17). This is contrasted with formative assessment by describing that any assessment activity can become formative if it 'provides information to be used as feedback by teachers, and by their students in assessing themselves and each other, to modify the teaching and learning activities in which they are engaged' (2003, p. 17). In order for formative assessment to work as a process it is vital that these activities cannot lie separate to the other themes of formative assessment, particularly those of feedback and peer- and self-assessment.

In this example, Wiliam (2011a) describes how an assessment activity can be used formatively rather than summatively:

> 'A middle school science teacher is designing a unit on pulleys and levers. Fourteen periods are allocated to the unit, but all the content is covered in the first eleven periods… the teacher gives the students a quiz and collects the papers. Instead of grading the papers, she reads through them carefully, and on the basis of what she discovers about what the class has and has not learned, she plans appropriate remedial activity for periods 13 and 14.'
>
> (Wiliam, 2011a, p. 41)

In some classrooms, tasks that elicit evidence of learning as part of formative assessment are often set to mimic higher-stakes summative assessment rather than designed as part of a wider process (Andersson and Palm, 2017; Birenbaum, 2015; Christodoulou, 2016; Coffey et al., 2011). This means that teachers may view the test itself as formative assessment rather than as part of a process (Popham, 2008). Wiliam claims that branding an assessment itself formative assessment 'just does not work' (Wiliam, 2011a, p. 39).

Alongside Black, Wiliam also claims that it is the use of the assessment rather than the nature of the assessment itself that makes it formative, explaining that the important aspect of it is the use of the evidence gained from the assessment by teachers and/or students and the decisions that come from the evidence gained. However, Wiliam also warns of a risk with such use of assessment in terms of the task design, stating that the information garnered from the assessment is only useful if the design of the assessment allows for strong instructional decisions to be made. He explains 'if the formative assessments are designed without any clear decision in mind, then there is a good chance that the information from the assessment will be useless.' (2011a, p. 45). This provides reason for teachers to consider both the design of assessment tasks, along with the how the evidenced garnered from them is used.

Since the publication of these teaching activities, Black and Wiliam have described a lack of connection between the principles of formative assessment and how the activities themselves are used in the classroom (Black and Wiliam, 2012). They call for a stronger theoretical foundation that would 'clarify both the connection of each of these types of activity to the central idea of formative assessment and might also indicate relationships between these kinds of activities' (p. 208).

The process of formative assessment relies on three main themes involving three types of person and has been more recently summarised by Figure 1.1.

<table>
<tr><td></td><td>Where the learner is going</td><td>Where the learner is now</td><td>How to get there</td></tr>
<tr><td>Teacher</td><td rowspan="3">Clarifying, sharing and understanding learning intentions and success criteria</td><td>Engineering effective discussions, tasks and activities that elicit evidence of learning</td><td>Providing feedback that moves learning forward</td></tr>
<tr><td>Peer</td><td colspan="2">Activating students as learning resources for one another</td></tr>
<tr><td>Learner</td><td colspan="2">Activating students as owners of their own learning</td></tr>
</table>

Figure 1.1

(Wiliam, 2011a, p. 46)

The table provides teachers with an overview of the process of formative assessment whilst also breaking down the stages that must take place into sections that teachers might focus on as units of this process. Black (2015) specifically highlights the need to develop informal assessments that enable evidence of learning to be sought. In his reflections on the implementation of formative assessment in schools, Black notes that many writers and teachers 'regard assessment as a peripheral component of pedagogy' (Black, 2015, p. 163) and explains that this may be due to weak links between assessment and pedagogy in historical literature. He proposes a model to bring the two together in order to assist teachers in improving the process of eliciting learning:

(1) Clear aims: The first stage of planning
(2) Planning activities: Setting up activities with the potential to achieve the aims.
(3) Implementation in the classroom
(4) Review of the learning: Using informal assessment to check achievement.
(5) Summing up: Using assessment to guide decisions about the next stage of students' work.

(Black, 2015, p. 164)

Black notes that ‘stage 4 is the one which deserves more prominence than it is often given’ (Black, 2015, p. 164). Black explains that the timing of such a review of learning is critical so as to allow for further correction and instruction if necessary. It is also important that the design of any activity be carefully thought through (Pellegrino, 2014; Popham, 2008). Stage 4 should also be used to assist students to ‘achieve an overview’ (Black, 2015, p. 164) of the work as a whole. This correlates with the ‘engineering effective discussions, tasks and activities that elicit evidence of learning’ section of Figure 1.

2.1.4 The tools for the process - activities that elicit evidence of learning

Whilst teachers may claim any particular task, activity or test as part of formative assessment, it is in the use of the activity that it actually becomes part of the process (Bell and Cowie, 2001; Bennett, 2015; Black 2003; Coffey, 2011; Gardner et al., 2010). The decisions that teachers and students make following assessment activities are crucial in making formative assessment successful. Wiliam explains the importance of task design and warns that ‘if the formative assessments are designed without any clear decision in mind, then there is a good chance that the information from the assessment will be useless’ (Wiliam, 2011a, p. 45). He suggests that assessment tools should be designed ‘backward’ so that the decisions that will be made from them are at the forefront of the design. There is previous evidence that teachers at School X (Author, 2017) regularly use exam-style questions as a tool for formative assessment, but the risk in this is that they turn in to simply ‘a series of 'mini' assessments each of which is essentially summative in character’ (Harlen and James, 1997, p. 365). These tests often provide an indication of which students are performing poorly but rarely provide accurate feedback to the student on where they need to improve (Heritage, 2018). A warning about using assessments designed for one use but then being used for another is given by Pellegrino, Chudowsky and Glaser who explain that ‘often a single assessment is used of

multiple purposes, in general, however, the more purposes a single assessment aims to serve, the more each purpose will be compromised' (2001, p. 41). This casts doubt over what the most effective tools for the process are.

It is common to find literature that suggests tasks and activities through which this process may be borne (Wiliam and Thompson, 2017), but the focus is often on short-term formative assessment, taking place in the minutes after teaching has taken place or at least in the same lesson (Black; 2015). Wiliam identifies the categories of short-, medium- and long-cycle formative tasks that range from minute-by-minute to across an entire year of education or a key-stage (Wiliam, 2006). Medium-cycle formative assessment can be seen as taking place within a unit of work, for example between one and four weeks.

Tasks that are designed to take place during, or shortly after, learning material has been introduced, are identified by Stobart (2008) as 'interactive' formative assessment. Alternatively, an activity taking place following a sequence of learning would be coined as 'retroactive'. Retroactive formative assessment tasks are often not the focus of literature and teacher training materials and the lack of development in assessment design (both formative and summative) amongst teachers is well-documented (Wiliam and Thompson, 2007; Bennett, 2010; Brown, 2004; Gardner, 2010; Koretz, 2008). This poses an opportunity to develop better retroactive formative assessment tasks. Stobart also refers to 'proactive' formative assessment which is classified as an activity that leads to modification to teaching following it. If an activity was retroactive, in that it took place after a teaching episode, but was then used to inform future activity, it may be said to be proactive.

In order to design effective tools for medium-cycle formative assessment so that they are able to benefit learning as much as possible, tasks must be designed for a single purpose. The purpose must be to facilitate a proactive process of formative assessment, rather than serve as imitation of other, higher-stakes assessment processes. Shepard, Penuel and Pellegrino (2018a) argue that this is best done by teachers:

> 'Instead of trying to make large-scale assessment models instructionally relevant, we argue that horizontally coherent curriculum-instruction-assessment-professional-development systems should be built from the bottom up, by teachers and subject-matter experts.'
>
> (Shepard, Penuel and Pellegrino 2018a, p. 52)

To create assessment systems in this way would mean that learning, rather than summative measurement, is placed at the centre of design thus making it more effective in informing learning rather than 'standing outside it and checking what has been learned' (Stobart, 2008, p. 146).

2.2 Linking assessment and learning

2.2.1 Reasons to link assessment and learning

'Our assessment tools constrain and shape the metaphors of learning with which we operate'

(Baird et al, 2017, p. 320)

In an education system that relies heavily on summative testing to make judgements of its efficacy and 'its use in new, neoliberal public management systems and structures' (Baird et al., 2014a, p. 3) it is not surprising to find that assessment holds a heavy influence over teaching and learning in classrooms. This assessment 'washback' (Alderson and Wall, 1993) means that there is a risk that any assessment activity becomes primarily about measurement. This measurement becomes the main focus of teachers and school leaders 'because the assessment results and the stakes attached to them can come to be seen as more important than the entity that they signify: the learning' (Baird et. al, 2014a, p. 3).

Heritage argues that the culture of high-stakes testing in schools has led to the warped use of assessment within classrooms, she comments that 'the replication of large-scale "testing formats" in classroom contexts with a corresponding distortion of instruction' may, alongside not benefitting learning as much as they might, also have 'detrimental effects on student identity development and motivation' (Heritage, 2018, p. 39).

An idealised relationship between assessment and learning is portrayed below, where there is a more reciprocal relationship between the two.

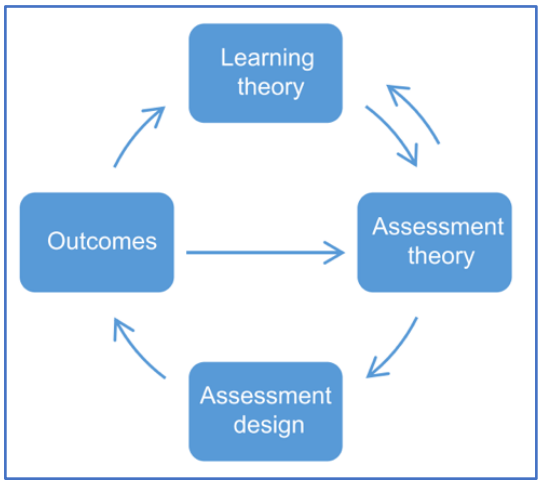


Figure 1.2

(Baird et al., 2017, p. 318)

The idea of assessment aiding learning is central to all literature on formative assessment but a purposeful link between formative assessment and theories of learning has not always been present in the literature (Wiliam, 2017). Baird et al. (2017) conclude that the relationship is difficult, explaining that, in the case of formative assessment, different strategies can be associated with different theories of learning and so agree with Bennett (2011) that formative assessment processes may be weakened by a lack of theoretical grounding.

Wiliam notes that 'over the last 20 years or so… there has been increasing scholarly interest in exploring more systemically the links between learning and assessment' (Wiliam, 2018, p. 42). Knowledge about the nature of learning is developing alongside this interest to integrate the two and Figure 1.2 suggests that a stronger relationship between theories of learning and assessment may then contribute to better assessment design (Baird et al., 2017; Shepherd et al., 2018a).

2.2.2 Learning theories

Theories of learning are most commonly presented under one of various headlines; behaviourist, constructivist, cognitive, sociocognitive or social constructivist and sociocultural theory. It is possible to view these perspectives not as distinct from one another but rather as successions, as learning theory encompasses increasingly complex ideas (Shepard, Penuel and Pellegrino, 2018b). Behaviourism is primarily concerned with observable stimulus-response behaviours and is cited as the most influential of all the learning theories on the nature of testing in education (James, 2006; Pellegrino, Chudowsky and Glaser, 2001). Cognitive theories of learning go beyond the observable stimulus and response mechanism of behaviourism and focus of the individual sense making occurring in the learner. Sociocognitive and sociocultural theories move again further; away from the individual to consider the role that social, historical and cultural factors have on learning. Shepard, Penuel and Pellegrino (2018b) treat these big theories flexibly to explain that they each encompass smaller theories that 'inform their development' (p. 23) and that there are aspects of all theories that can be useful in informing assessment practices. For example, whilst sociocultural theories of learning are supported by many in recent literature (Wiliam 2018), Shepherd et al. explain that 'the sense-making focus of cognitive models is still a valued aspect of contemporary learning theory' (Shepard, Penuel and Pellegrino, 2018b, p.23).

James explains that teaching and assessment practices do not always 'take sufficient account of current understanding of the ways in which students learn subject matter' (2006, p. 47). This provides an opportunity to develop practices that more effectively combine learning theory with assessment practice. Pellegrino, Chudowsky and Glaser describe a model of learning as an essential component of teaching, describing it as a 'nucleus that brings cohesion to curriculum, instruction, and assessment' (2001, p. 3).

2.3 Using cognitive science to understand how learning takes place

With curriculum time scarce, it is important to make good use of it and so it is beneficial, and perhaps necessary, that any regular assessment activity should be designed to not only serve a purpose for future learning but also aid learning in itself. The field of learning sciences has become increasingly prominent in educational thinking (Wiliam, 2018; Heritage, 2018) and research from the field can be used to consider both the design and the placement of assessment tasks in the curriculum, as well as inform teachers in making conclusions and decisions that are as valid as possible. In addition to helping teachers think about the design of an assessment task, the learning sciences also provide insight into the nature of learning itself.

2.3.1 Learning and memory

Learning and memory are intrinsically linked, and the mechanisms at play within a person's brain are highly complex with many aspects still opaque to specialists (Baddeley, Eysenck and Anderson, 2009). In addition to this complexity, it is important that teachers exercise caution and stoicism when looking to fields such as neuroscience to improve education. There are limits and perhaps even disadvantages to applying brain research to environments outside of

strictly controlled laboratories (Blakemore and Bunge, 2012; Bruer, 1997; Skinner, 1985; Willingham, 2008; Willingham, 2009). Despite this, there are opportunities offered by research in the fields of learning sciences that may benefit teaching and learning in schools (Dubinsky, Roehrig and Varma, 2017; Willingham, 2008).

2.3.2 The testing effect

Whilst there is a danger of large-scale testing practices being replicated in classrooms, reducing formative assessment practices to little more than a data collection exercise, there is a body evidence that shows testing to be beneficial to learning.

The testing effect, or retrieval practice as it is also known, is an old concept (there is evidence that Aristotle described it in his work) that is increasingly supported by research (Einstein, Mullet and Harrison, 2012; Karpicke, Butler and Roediger, 2008; Littrell-Baez et al., 2015). Karpicke and Roediger describe how the prevailing attitude to testing in the fields of memory research and educational practice has been one that depicts testing as a 'neutral event that merely measures learning' (2008, p. 966). Instead of only measuring learning, the authors explain that testing improves 'long-term retention' much more effectively than other strategies such as 'repeated study' (p. 967). This would suggest that assessments that require the student to recall from memory unaided may benefit the student more than re-study time or if the assessment task was undertaken with prompt material (for example textbooks or classroom notes).

A use of testing in this way may be labelled negatively by some as 'drilling' and prescribing to an outdated, Dickensian view of education (Christodoulou, 2014). However, if the tasks can be designed to use this evidence to the effect of improving learning whilst simultaneously

allowing for diagnosis of learning as part of formative assessment, it may provide promise. A criticism of this type of activity is that students are simply taught to memorise facts rather than be able to transfer and apply knowledge. In a study that contradicts this view, McDaniel et al. found that 'a spaced quizzing procedure with feedback can promote learning that is deeper than just retraining a particular answer' (2013, p. 368). In one experiment, the quiz items simply required students to define key terms but the authors found an improvement in transfer. Alongside these findings the authors stress the importance of other factors such as repeated and spaced presentation of content along with feedback, but the research does counter the view that a focus on knowledge and testing may in some way make a student a less flexible learner and adds a dimension of use beyond assessment tasks as purely diagnostic. This small-scale study is one in a larger body of evidence supporting this effect (Einstein, Mullet and Harrison, 2012; Karpicke, Butler and Roediger, 2008; McDaniel et al., 2011; McDaniel et al., 2013; Roediger and Butler, 2011).

The idea that a task that closely resembles a traditional test (in that students are in exam conditions, retrieving from memory) may be best placed as a tool for formative assessment is not always supported by the literature. For example, James warns against the use of 'short answer tests in science that require recall of facts' (2006, p. 47) as she explains that it is also important to assess a student's ability to investigate. However, the research on the testing effect suggests that there may be a role to play for the recall of facts to enhance learning.

2.3.3 Learning versus performance

Kirschner, Sweller and Clark claim that the aim of all educational instruction is 'to alter long-term memory' stating that 'if nothing has changed in long-term memory, nothing has been learned' (Kirschner, Sweller and Clark, 2015, p. 77). This definition of learning is debatable but

poses important questions for the timing of assessment tasks so that the interpretations made by teachers and students are as valid as possible. A focus on short-cycle formative assessment activities in the literature on formative assessment may lead teachers into thinking that evidence presented by a student demonstrates this long-term learning. Wiliam addresses this risk and justifies the focus on short-cycle formative assessment in stating that 'whilst we cannot be sure students have learned something because they know it at the end of the lesson, if they do not know something by the end of the lesson, it is unlikely that they will know it weeks later' (Wiliam, 2016, p. 124); but it is important that teachers appreciate this distinction (Pellegrino et al., 2001, p.3). Soderstrom and Bjork (2015) define this difference as 'learning versus performance' (p. 176). They describe learning as being a longer-term process than performance and stress the need for awareness of the difference between the two. They describe performance as a 'temporary fluctuation in behavior (sic) or knowledge that can be observed and measured during or immediately after the acquisition process' (2015, p. 176). The risks posed to teachers from this idea is that in their assessment of learning they instead assess short-term performance from a student that may not be repeatable even a day later.

An alternative way of thinking about this idea is through the use of the terms *storage strength* and *retrieval strength* (Bjork and Bjork, 1992) where retrieval strength represents 'the current ease of access or activation of these memory representations' (Soderstrom and Bjork, 2015, p. 191) and storage strength relates to 'the degree to which memory representations (i.e. knowledge and procedures) are integrated or entrenched with other memory representations' (Soderstrom and Bjork, 2015, p. 191). If formative assessment is to aid learning, then tasks should enable students to increase storage strength in their memory while maintaining high retrieval strength. Soderstrom and Bjork also describe how the two

are linked; they suggest that the higher the retrieval strength of a representation in memory, the lower the gain in storage strength upon retrieval (2015). There are two implications of this idea for classroom teachers. First, that distributed practice may be more beneficial than massed practice, so as to allow for an amount of forgetting to take place before an idea is re-visited. Secondly, that tasks requiring more effortful, un-aided retrieval may provide a larger improvement to learning than tasks that reduce retrieval effort (for example, allowing students access to resources when answering questions rather than requiring them to answer them form memory). Consequently, successfully making inferences about learning as part of formative assessment is a complex process and raises questions about the timing of assessment events.

2.4 Improving the validity of teacher and student judgements of learning

In his criticisms of the implementation of formative assessment Bennett (2011) describes a lack of accuracy in the judgements that teachers and students make about learning (the measurement issue). One way in which the degree of inaccuracy might be reduced is through viewing learning and performance separately and by teachers placing assessment activities at a suitable distance away from the initial classroom introduction of the topic. This would aim to improve the likelihood that long-term memory is being tested and may also hold a second benefit in increasing storage strength at a time when retrieval strength has been reduced (Soderstrom and Bjork, 2015). Another aspect that needs consideration is that of task design. Teachers are often involved in the creation of assessment tasks used in the formative assessment process but there is a lack of education for teachers as part of both their initial training and subsequent development in schools (see 2.1.4).

2.4.1 Defining validity

Having established that the design of any assessment activity is integral to its eventual use, Pellegrino, DiBello and Goldman note that 'relatively little attention has been given to discussion of the design and validation of such assessments' (2016, p. 59). The authors propose a conceptual framework aiming to support teaching, learning and assessment that is made up of three components:

- Cognitive validity
- Instructional validity
- Inferential validity (p. 59)

Pellegrino, DiBello and Goldman acknowledge the documented rise in influence that psychology and other learning sciences have had on educational assessment and link this to the original claim by Black and Wiliam that 'any assessment must be optimized for its intended purpose and interpretive use' (p. 59). The authors also seek to consider the role that empirical evidence and theories of cognition and learning may have in 'guiding and constraining assessment development' (p. 60). The conceptual framework that they propose will be examined and the framework will continue to be used in the discussion that follows this section.

Cognitive Validity

Pellegrino, DiBello and Goldman explain that for an assessment to have cognitive validity it must be designed so that knowledge and skill can be assessed without another 'aspect of cognition such as language or working memory load' (2016, p. 62) impeding the assessment. The authors state that it is important to consider how knowledge of a curriculum area builds in a student over time and that the assessment aims to access evidence of this learning at an appropriate point in a students' learning (Pellegrino, DiBello and Goldman, 2016).

Instructional Validity

Instructional validity may be achieved if an assessment is 'alligned with curriculum and instruction' (Pellegrino, DiBello and Goldman, 2016, p. 62). In practice, this requires assessment designers to pay close attention to what teaching has occurred and which curriculum has been followed in order to then design an assessment which reflects this as genuinely as possible. The difference between cognitive and instructional validity is that cognitive validity is more concerned with access to a question or task by the student and instructional validity ensures that once accessed, the question or task assesses something that the student should have received instruction in.

Inferential validity

An assessment to be used for diagnostic purposes should be judged on the 'extent to which an assessment reliably and accurately yields model-based information about student performance' (Pellegrino, DiBello and Goldman, 2016, p. 62). The nature of inferential validity means that it sits alongside both instructional and cognitive validity but extends the scope of the assessment further in terms of its formative use so that teachers and students can accurately diagnose progress towards learning aims.

The authors state that there should be constant reflection on the validity of an assessment during all stages of its development and that this should continue as it used in educational settings.

2.4.2 Constructing test items

In order to be more confident in evidence of learning rather than performance, students should be required to respond to tasks with a time distance between the two (Dunlosky et al., 2013). In addition, the testing effect suggests that the biggest learning gain is gained from unsupported retrieval. These factors support the idea that effective formative assessment activities may take a similar basic form to a more traditional summative test (often called an end-of-topic test) in that it asks questions of students under test conditions.

Two styles of questions available to teachers are selected-response items and constructed-response items. Both come with significant advantages and disadvantages (Osterlind, 1998), but the scope of this review is to determine which would be more suitable to fulfil a diagnostic function for both teacher and student. Selected-response items may include 'binary-choice items, matching items and multiple-choice items' (Popham, 2003, p. 72), they are able to cover a lot of content very quickly and are easy to mark as either correct or wrong for both teacher and student. Popham notes that binary-choice items are fallible to promoting simple memorisation, especially if the questions are not well designed (2003, p. 74).

Constructed-response items can take two forms: 'short-answer items and essay items' (2003, p. 86), Popham is quick to declare that 'the chief virtue or any type of constructed-response item is that is require students to create their responses rather than select a pre-packaged response from the answer shelf' (p. 86). Popham describes the main disadvantages in terms of the time requirement of both student and teacher; written responses take longer for the student to construct thus potentially reducing the scope of the assessment across the learnt domain and they are also more laborious and more difficult for teachers to mark. He also

notes that if questions are not structured enough they may fail to elicit any evidence from the learner (2003, p. 89), giving a reduced level of cognitive validity to judgements made by students and their teachers (see 2.4.1).

2.5 Self-regulation

Cognitive perspectives of learning are primarily concerned with thinking about how learning takes place with an individual. However, to take heed from a more sociocultural perspective, the learner cannot be viewed as a neutral entity within a sterile environment and other, social factors may place a key role in learning success (Shepard, Penuel and Pellegrino, 2018b).

2.5.1 What is self-regulation?

It has become increasingly common to use the terms metacognition and self-regulation as desirable outcomes of education in the UK. The Education Endowment Foundation report that the section on metacognition and self-regulation in their toolkit is one of the most highly viewed, having been accessed online more than 120,00 times (Education Endowment Foundation, 2018, p. 4). There is also evidence to suggest that it is an effective way to boost achievement. Hattie reports an effect size of 0.69 (Hattie, 2012, p. 267) for metacognitive strategies and places it at number fourteen in a list of one hundred and fifty influences on achievement in education. A recent guidance document produced by the EEF identifies three aspects of self-regulation; cognition, metacognition and motivation (2018 p. 9)

Zimmerman defines self-regulation as 'the self-directive process by which learners transform their mental abilities into academic skills' (2002, p. 65). The process of self-regulation may be broken down into three phases: forethought; performance and self-reflection. These stages

may also be viewed as the stages before, during and after efforts to learn have taken place (Zimmerman, 2002). The forethought phase requires students to analyse tasks prior to completion but consideration is also given to the self-motivation for the task. During the performance phase self-control and self-observation are key. Self-control means that a student is able to implement the strategies considered during the forethought phase but only with the right motivation. Self-reflection includes the evaluation of one's performance in a task and self-reaction occurs following this where it can take positive or negative meaning.

Successful self-regulation also relies on student being able to rely on sound learning techniques so that they can develop learning behaviours that see them make good progress in their learning. In their research, Dunlosky et al. (2013) identify two techniques as having particularly high utility; practice testing and distributed practice. If habits in these practices can be developed through regular formative assessment opportunities, then they may serve to improve learning in both the short and long-term.

It is important to develop assessment alongside an appreciation for developing motivation and self-regulation in students (Baas et al., 2015; Clark, 2012; Karpicke, Butler and Roediger, 2011). Involving students at all stages in this process is a feature which aligns wholly with the principles of formative assessment (Wiliam, 2011a), as well as designing assessment tasks and opportunities that students can then use in the future to improve their learning. Nicol and Macfarlane-Dick explain that involvement of the student in formative assessment shows students what 'good performance is' and 'facilitates the development of self-assessment' (Nicol and Macfarlane-Dick, 2005, p. 7).

Chapter 3

Methodology

3.1 Research Questions

The intervention implemented across the science faculty at School X draws theoretical ideas from a number of different areas: the widely referenced body of assessment for learning literature; effective learning as proposed by cognitive science; and assessment design. However, all of the literature themes centre around one idea, to improve the learning of all students over time and as such the research questions are purposefully broad so that they are able to cover sufficient scope for what data may arise. The following research questions form the frame for data collection:

Research Question One

What impact does the assessment intervention, designed to improve learning over time, have on summative grades for students in years 7, 8, 9 and 10?

Research Question Two

What are students' views of the impact of the assessment intervention on their learning over time?

Research Question Three

What are teachers' views of the impact of the assessment intervention on their students' learning over time?

Approach

This practitioner research can be viewed as spanning both traditional positivist and interpretivist research paradigms (Denscombe, 2017, p. 8). With both quantitative and qualitative data being sought and analysed in order to address the research questions, this research can be viewed as mixed methods research (Johnson, Onwuegbuzie and Turner, 2007). A mixed methods approach to research allows for different methods to be used so that the issue may be analysed from a variety of different viewpoints. These multiple viewpoints then allow for triangulation and a more complete consideration of the issue (Denscombe, 2017). In combining these approaches, the project can be identified within a pragmatic grounding where the choices of method seek 'to meet the needs of the research rather than the allegiances or preferences of the researcher' (Cohen et. al, 2018, p. 34). Furthermore, because it is difficult to measure learning, partly because it cannot be easily defined, a 'convergent parallel design' (Creswell and Plano Clark, 2011, p.77) approach is taken. This means that the research questions not only seek to evaluate the impact on learning from multiple perspectives, but these perspectives are also used to inform one another.

3.2 The Intervention: Progress Checks

Previous school research had shown that there was a variation in how teachers viewed and conducted formative assessment, including within the science department (Author, 2017). The research showed that teachers were more confident in the day-to-day mechanics of formative assessment than they were in embedding it over time, across a unit of work, a term or even a key stage. This somewhat hazy vision of what good formative assessment looks like was further muddled by the term 'diagnostic' being attributed to summative tests that in reality served very little diagnostic function. The focus of formative assessment as being

central to the learning process going forward had seemingly been usurped by the habit of assessing learning at the end of a unit, after all of the lesson time had been used.

It was decided amongst the faculty leadership team that, in light of the research literature, we would re-design the assessment tasks that students took at the end of each topic. Alongside the change in the tasks themselves, the timing of the task within a unit of work was also modified.

The main reasons for discarding the previous system of end of topic tests were:

- The end of topic tests that we had previously used were made up of exam style questions designed to contribute to a summative judgement.
- The coverage of the test across the learnt domain was low which meant that sometimes tests missed entire topics within a unit.
- There was an over-reliance on the tests in terms of diagnosing problems with understanding, but the problem areas were predictably those which featured in the hardest question in the test; a poor judgement in terms of its validity.
- Marking of the tests mainly comprised of ticking and crossing alongside a percentage or grade.
- In some classes the tests were reviewed but with no curriculum time dedicated to a review period this was inconsistent.

Progress checks were implemented across key stages three and four and, in many respects, directly replaced end of topic tests. One another change occurred at the same time where curriculum time was re-shuffled in order to build in at least one (usually two or three) **review lessons** following the progress checks in order for the teacher to feedback and for the students to reflect on their areas of strength and weakness and to improve in these.

The key features of the progress check tasks are summarised below, with reference to the literature that support the design choice.

The task was sat part-way through the unit, not at the end

This recommendation is made by Black (2015) when describing the importance of affording opportunity for improvement as part of formative assessment. In addition to reflection, the timing of the tasks in this way means that there are multiple points in the curriculum in which students will encounter the ideas that they need to learn. Firstly, in initial learning then again during the progress check and then a third time in review if necessary (Nuthall, 2007). The timing of the tasks also allows opportunity for students to forget before practicing retrieval. The literature suggests that this may improve the insight into a student's learning (rather than performance) (Soderstrom and Bjork, 2015; Bjork and Bjork, 1992) as well as having a larger impact on improving storage strength of the content.

Feedback centres around curriculum points and students do not receive a grade

Some of the original research undertaken by Black et al. (2003) reported the detrimental impact of grades of the aims of formative assessment. In addition to this research, the findings of a recent review into written marking suggests that coded feedback can be just as useful as written comments (Elliot et al., 2016).

There is a style of open questioning, with no multiple choice.

Popham (2003) explains that the benefit of open questions is the ability of them to expose a students' thinking. Alongside this insight, open questions do not allow students to guess the correct answer, this improves the cognitive and inferential validity (Pellegrino, DiBello and Goldman, 2016) of judgements made from the progress checks.

The questions are focused on the core knowledge required of them.

There is also a body of research that suggests, perhaps counter-intuitively, that

that a focus on definitional items can improve transfer in later tests (Liu et al., 2017; McDaniel, 2013; McDaniel, Roediger and McDermott, 2007). Furthermore, questions that can be re-used can form part of a students' revision, encouraging the use of the testing effect (Agarwal, Bain and Chamberlain, 2012; Roediger and Butler, 2011) and self-regulation (Bjork, Dunlowsky and Kornell, 2013; Karpicke, Butler and Roediger, 2009; Zimmerman, 2000)

Students were given feedback immediately in the next lesson

The importance of formative assessment to be a process rather than an instrument is emphasised by many (Black, 2015; Coffey et al., 2011; Harlen and James, 1997; Sadler; 1998; Torrance, 2007). The immediate use of the tasks to inform future learning ensure that this reflection is a key part of the process thus allowing students to understand 'how to get there' (Wiliam, 2011a) as quickly as possible.

The tasks are sat in exam conditions

The testing effect states that the largest learning benefits come from un-aided retrieval by an individual (Agarwal, Bain and Chamberlain, 2012; Roediger and Butler, 2011).

3.3 Data Collection

Research Question	Research Instruments	Method of Data Analysis
What impact does the assessment intervention, designed to improve learning over time, have on summative grades for students in years 7, 8, 9 and 10?	Collection of scores attained in end-of-year summative exams (June 2018) sat by students in years 7-10.	Raw scores are compared with the previous cohorts' scores on the same test using an unpaired T-test to establish difference, and then any level of significance of this difference.
What are students' views of the impact of the assessment intervention on their learning over time?	Student surveys were carried out in across years 7-10. There were 506 surveys collected in total. Individual interviews with six students in Year 11.	Quantitative analysis of survey items to indicate trends across the cohort (Appendix D). Content analysis of open response questions using analytic coding (Cohen et. al, 2018) (Appendix C). Analytic, followed by axial coding of interview transcripts (Cohen et. al, 2018) (see Appendix F).
What are teachers' views of the impact of the assessment intervention on their students' learning over time?	Individual interviews with six science staff.	Analytic followed by axial coding of interview transcripts (Cohen et. al, 2018) (see Appendix F).

Data collection methods for investigating Question One

What impact does the assessment intervention, designed to improve learning over time, have on summative grades for students in years 7, 8, 9 and 10?

Summative exam data

The data collected for this question is quantitative and consists of student performances in summative examinations taken at the end of the school year which are then compared to the previous years' data using the same exam papers.

An unpaired t-test was conducted on all data in order to conclude whether or not there was a statistically significant difference between the year on year results. The choice to use an unpaired t-test was necessary as the two groups of students sitting the same tests one year after the other comprised of different students. This sort of t-test is less effective than a paired t-test (Cohen et al., 2018) but this wasn't a possible measure to make in a school and across different year groups. The assumptions made in an un-paired t-test include the following:

- *Population is normally distributed*
- *Parametric data*
- *Each group is independent of the other*
- *The groups to be compared are nominal*
- *The sets of scores have approximately equal variances, or the square of the standard deviation is known*
- *The data are interval or ratio*

(Cohen et. al, 2018, p. 845)

It is common in education to use effect size calculations when making quantitative judgements however the lack of a control group in this instance meant that this would not have been suitable.

Moderation of all exam marking takes place within the faculty after each exam series which improves the reliability of the data used. The main issues with any conclusion drawn from this data are:

1. Different year groups are taking the tests each year
2. Different teachers are teaching the curriculum

These two factors were considered when designing the project. A comparison of year groups was conducted in order to deduce whether or not it was expected, due to prior data, that one year group would perform better than another (Appendix G). The data show that years 7-9 are largely similar in these terms but there is a noticeable difference in the higher targets of students in Year 10 which will need to be considered in the analysis of summative assessment data. It is expected that different teachers will teach different groups in each year and indeed it is important, particularly for the development of newer teachers, that they experience this variation in order to develop their expertise and gain a sound knowledge of the secondary science curriculum. This wider priority surpassed the desire for consistency in this area for the project. Despite these changes, the teaching staff across the faculty did remain constant over the two years so they were all in a position to make comparisons.

The decision to include this form of evidence is also justified by a lack of empirical evidence in the literature (Baird et al., 2014). Whilst there are extensive limitations in the ability of this project to contribute to this field meaningfully, with a lack of control group alongside many extraneous variables and non-standardised measurement instrumentation, it does provide an alternative perspective through which to make observations and reflect on the efficacy of the intervention.

Data collection methods for investigating Question Two

What are students' views of the impact of the assessment intervention on their learning over time?

Student survey data

Students across all year groups were surveyed in June 2018. The collection of the data was timed so as to allow students exposure to the new practice for as long as possible but to also allow for data to be analysed in time for developing practice for the following academic year. To allow for maximum levels of response, these were completed on paper and in lesson time.

The questions in the survey are closely aligned with the interview questions so that they can provide useful triangulation upon analysis. The much larger sample size of the survey also provides a measure of reliability to conclusions that match with the interview data. However, whilst able to cover a large number of students, the complex nature of the topic is not best served by data collection in this way. The main purpose of the student surveys is to gain, in an efficient and timely manner (Cohen et. al, 2018), an insight into basic levels of confidence in the intervention to improve learning and to identify, from a student perspective, the most helpful and the least helpful aspects of the intervention. This data provides a cohort-level view of different year groups as well as from the student body as a whole.

Question Design

The question design for this section of the survey potentially leads the students into a positive response due to the phrasing of the questions being positive themselves. In an attempt to reduce the potentially leading nature of the question, the 'strongly disagree' category was

placed at the left-hand side of the scale. Research has shown respondents to be biased towards the left-hand side of survey items (Hartley and Betts, 2010, p.25) so this was a choice that attempted to balance the potential bias within the statement itself. Question 4 was included as a 'check' question (Denscombe, 2017, p. 193). The level to which students think the intervention has benefitted their learning is of high importance to the research questions and so a second question, differently worded but with similar meaning, allowed for a more reliable conclusion to this question.

Student interviews

Six Year 11 students were interviewed in May before they started their GCSE examinations. I selected a Year 11 group that I had taught for the previous two years to conduct interviews with. Within this group, I pre-selected a group that was representational of the group as a whole in terms of prior attainment, sex, ethnicity, special educational needs, English language proficiency and whether or not the school receives pupil premium funding for the student. I purposefully chose a group that I had taught so that I could be sure of their experience with the progress checks and decided that I would be in a better position to make inferences from their responses. I also knew that this class were quite vocal in discussing their views about their learning with me, something that I hoped would continue across into interviews, reducing the impact of hesitancy and wariness that may have occurred if I was to of interviewed students from other classes.

Sampling considerations

In the analysis of this data it is important to note that the students were selected from a separate science Year 11 chemistry class that I have taught for two years. Most of the students are amongst the most able in the year group and they are a well-motivated group of students. The benefits of using this class were:

1. A good level of trust exists between me and the students and they appear happy to be honest in their views with me; whilst teaching them I have experienced them being both honest in terms of positivity and critique.
2. When interviewed, the class were in the summer term of their GCSE year, likely to be highly engaged and thoughtful about their learning with a good experience of how assessment has changed across science in the last two years.
3. Frequent contact in lesson time made scheduling (and re-scheduling) convenient.
4. Having taught the class I had a reasonable level of assurance that they have experienced the progress checks in the manner in which they were intended.

The following disadvantages must be considered:

1. No matter how honest the students appear to me, a degree of acquiescence (Cohen et al., 2018) must be considered likely; they knew that the research project was important to me and that I have put in large amounts of effort in the design of the progress check tasks.
2. The students interviewed are, by most common measures, considered able and successful in education so far and are thus more likely to describe the positives of any education system.

Interview Questions

Interview questions for both students and teachers were based around the same themes as the student survey (Appendix E). They primarily sought to illicit the ideas of students and teachers about positive and negative impacts to learning as well as suggestions for improvements in the future.

Data collection methods for investigating Question Three

What are teachers' views of the impact of the assessment intervention on their students' learning over time?

I conducted interviews with six full-time science teachers towards the end of the academic year. Teachers were emailed with details of the project and asked to opt-in to being interviewed at a time convenient for them.

It was important to emphasise that my position in the interview was that of researcher and not that of head of faculty although it cannot be claimed that this would not have been a factor in influencing the responses of my colleagues who may have felt under some pressure to be more positive about the intervention knowing that I initiated it and have put a large amount of work into creating the resources for it. The interviews were semi-structured with a skeleton list of pre-written interview questions that allowed for deviation and elaboration if necessary.

3.4 Ethical considerations

This research project has been conducted using the British Educational Research Association (BERA) Guidelines (2018) and approval from the Central University Research Ethics Committee (CUREC) was sought and gained prior to undertaking the study.

Written permission for the project was sought and granted from the Headteacher of the school (Appendix A; Appendix B). Denscombe (2017, p. 341) outlines four principles that should be followed by social researchers:

1. *Participants interests should be protected*

All references to individuals have been anonymised in all notes and transcripts. Interviews were recorded and stored on a password protected device and hard copies of evidence have been stored in a locked storage unit.

The decision for not having a dedicated control group within the project was also made on largely ethical grounds. The supporting literature gives reason to think that the intervention may benefit the learning of students. Alongside this, the considerable amount of work and effort that went in to creating the resources used in the project and the consequent time dedicated in faculty meetings to developing the intervention meant that denying a group of students the opportunity to partake in the intervention may have led to a reduced focus on them as a group.

2. *Participation should be voluntary and based on informed consent*

All participants in interviews were informed of the project aims by an information sheet before being asked to sign that they understood the aims of the project and that they could

withdraw their consent to participate at any time. Students completing the survey were told by their teachers about the project as well as it being detailed at the top of the survey, they were also informed that they could withdraw their participation at any time.

3. *Researchers should operate in an open and honest manner with respect to the investigation*

It was important for me to be open and honest with the science faculty from when we started that I would be undertaking analysis of changes that we were making to our formative assessment practices. When interviewing my colleagues, I emphasised the importance of honesty and specifically welcomed criticism of the intervention.

4. *Research should comply with the laws of the land*

All research was conducted within the laws of the land as well adhering to school policy.

3.5 Collaboration

The nature of this research project meant that all science teachers were deeply involved in the implementation of the intervention throughout the school year. Faculty meeting time, especially in the first term was used to review and refine the progress checks, this continues at the time of writing and will do so well into the next academic year. On a day-to-day basis the involvement of colleagues was in using the progress checks tasks with their classes as per the faculty plan but there was deeper involvement from the key stage three and key stage four coordinators who took on some of the writing of the tasks after we had established an initial format for them. In handing over some of the writing to the coordinators the style of questions and the coverage required was communicated and I then checked all progress checks before they were used by classes.

The discussions that have arisen between myself as head of faculty and my colleagues this year has also allowed us to develop the team's awareness of the theory supporting the intervention as well as develop their skills in assessment in their own classroom. The intervention has brought formative assessment to the forefront of faculty practice and I feel confident that, despite the predictably differing opinions on its efficacy and its integration into the classroom, the teachers I work with now have increasing levels of expertise in the implementation of assessment for learning from theory in to classroom practice.

The project and the findings from it have also been shared more widely; a group of four visiting science faculty leaders from partnership schools in London attended School X for an afternoon of presentation and discussions in June 2018 which has led to continued work and sharing of the progress check model. The project was also shared at a poster conference at the University of Oxford in June 2018 where some teachers expressed keen interest in hearing

more about the outcomes and there has since been email communication with these teachers.

The senior leadership team at School X has identified assessment as a focus for teacher professional development in the academic year 2018-19 and so the theory behind, and impact of, this project was presented to all staff in September 2018. This has led to discussions between me and the leaders of other faculties in the school interested in developing assessment practices in their faculty areas.

Chapter 4

Findings and Discussion

This section will begin with a presentation and review of the evidence gathered through each data collection method, followed by an analysis of the main research findings in relation to the literature.

4.1 Student surveys

Attitudinal insights

Questions 1, 3 and 4 were designed to assess the attitudes of students after they had been involved in the intervention for most of the academic year.

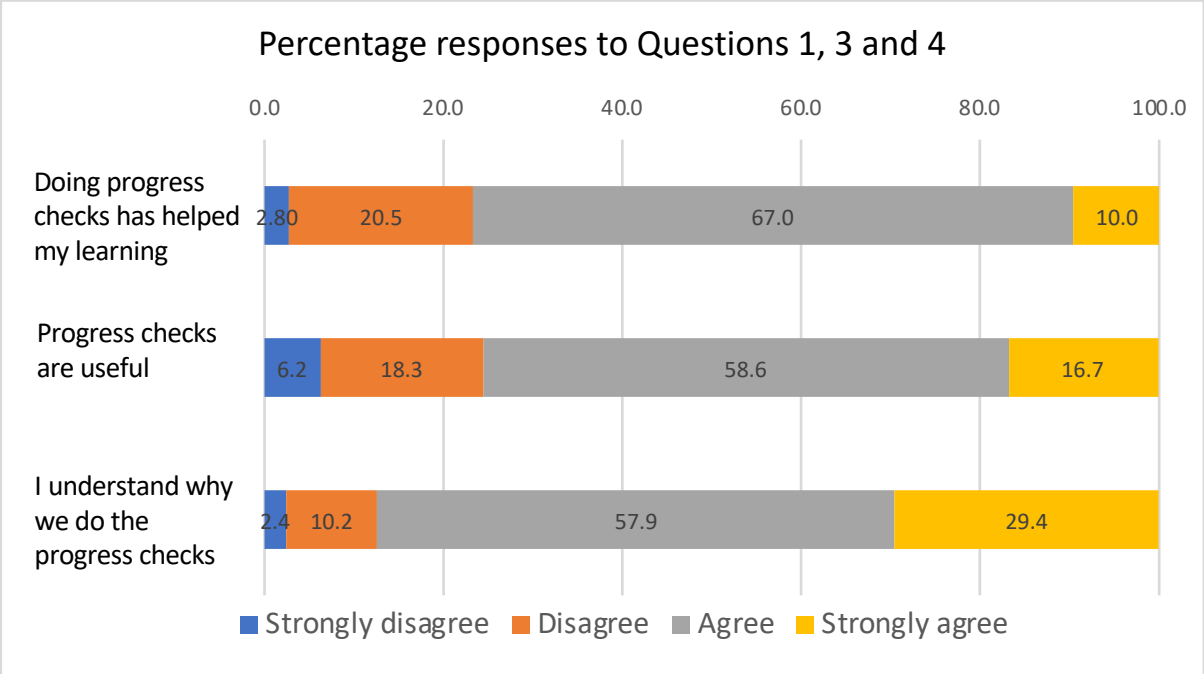


Figure 4.1

As Figure 4.1 shows, 77.0% of the students surveyed either agreed or strongly agreed that the progress checks have helped their learning. This positivity towards the intervention is mirrored in the check question regarding the usefulness of the intervention where 75.3% of students either agreed or strongly agreed with the statement. The third question in this series aimed to assess whether students saw and understood the purpose of the intervention; a

large majority of students' responses positively to this statement with a total of 87.3% students either agreeing or strongly agreeing.

The distribution of responses amongst students from different year groups was quite consistent with no question eliciting a significantly different response from one year group compared with another.

Function

The data gathered from the two questions shown in Figure 4.2 were designed to explore whether some of the intended functions of the intervention were recognised by the students.

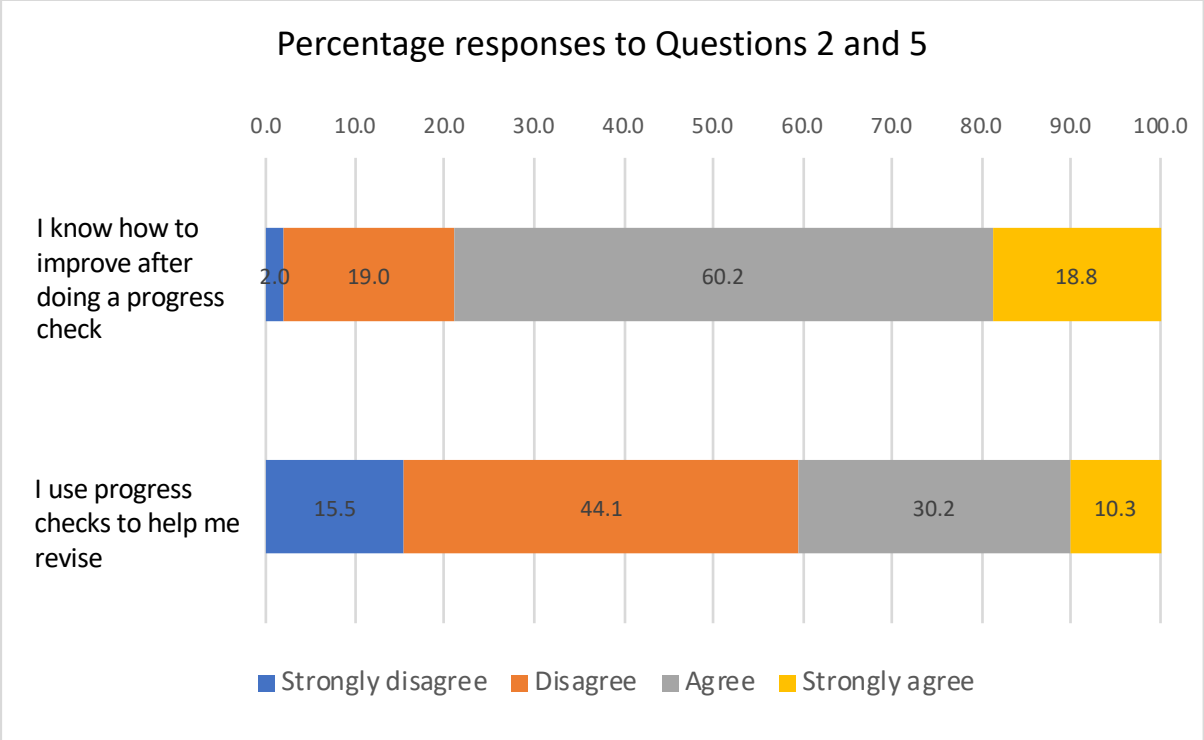


Figure 4.2

The data suggest a clear difference between the success of the intervention to provide feedback and its function as a revision tool. In total, 79.0% of students agreed or strongly agreed that they know how to improve following a progress check which suggests a degree

of success in this aspect of the design. The data suggest that the revision function has been less successful with the majority of students disagreeing or strongly disagreeing that they use their progress checks to revise. It could be argued that younger students may not yet be revising on a regular basis and so it would be unrealistic to expect them to reply positively to this question however, contrary to this, it is the youngest year group (Year 7) that reported the most agreement with this statement (47% agreed or strongly agreed) compared with the most senior year group (Year 10) where only 28.8% of the cohort agreed or strongly agreed with the statement. This question proved to be one of the most inconsistent across the different year groups and warrants further exploration.

Effect on summative attainment

This question only provides weak evidence for the third research question of this project but also serves as comparative data to the summative exam results. The majority of students (63.4%) agree that progress checks improve their exam performance. The difference between this and question one may suggest that, for students, there is a disconnection between exams and learning.

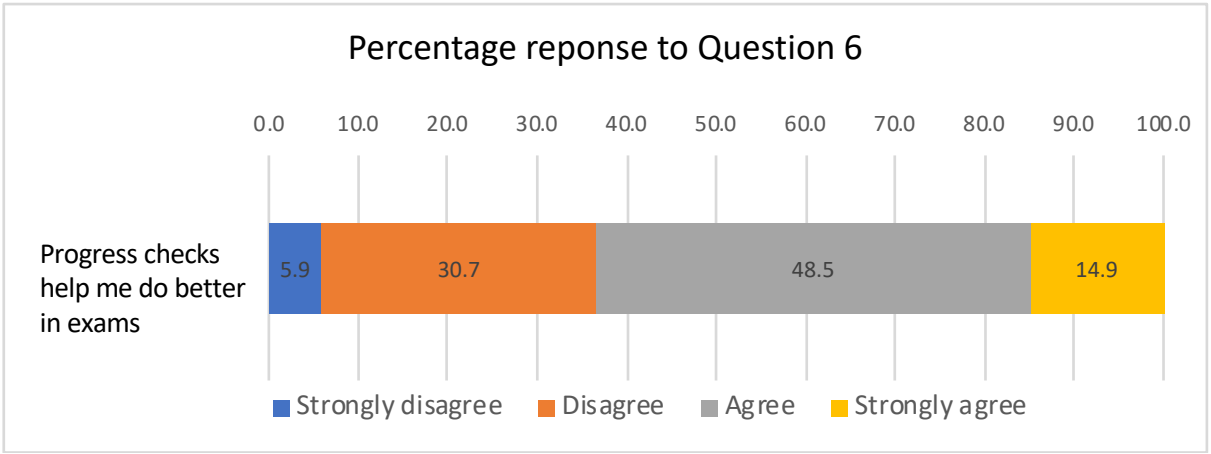


Figure 4.3

Comparison to pre-intervention

After analysing the comments made in the follow-up question to this survey item, it was decided that only Year 9 and Year 10 student responses would be considered for this question.

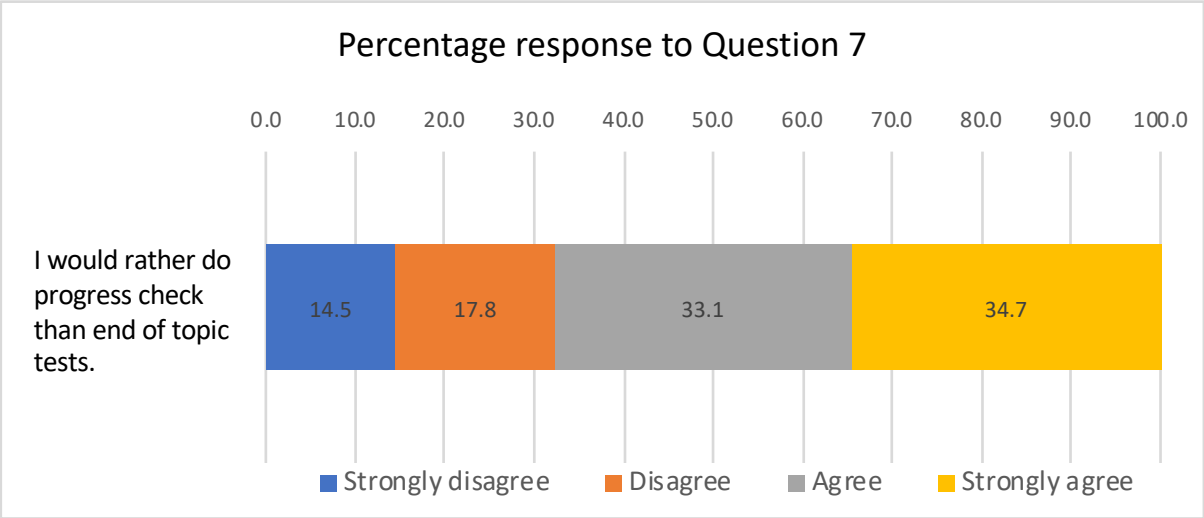


Figure 4.4

The question provides a basic overview of student opinion, which is further supplemented by the student interviews, and shows that 67.8% of students prefer the progress check tasks (the intervention) to end of topic tests (the style of task used pre-intervention). The spread of data suggests stronger feelings towards this question compared with the other question asked in this section of the survey with almost half of the students selecting either 'strongly disagree' or 'strongly agree'. In designing the survey items, this stronger feeling was pre-empted from my own experience throughout the year which gave reason to allow for an open response explanation for students following their ranking of this statement.

Survey Question 8

Question: Please explain your answer to the last question.
(The question prior to this was: 'I would rather do progress checks than end of topic tests')

After feedback from colleagues and sampling of the responses for this particular question, it became clear that students in Year 7 (who have never done end of topic tests in science) were very unsure of their responses and those in Year 8 appeared to be equally confused. For this reason, the answers from these students were removed from analysis, leaving only the responses of Year 9 and Year 10 students.

Agreed with a clear reason identified	Total Number	Percentage
C - Less stressful	32	28%
I - Easier to identify improvement	21	19%
D - Better for revision	16	14%
J - Topic coverage	15	13%
A - Easier	14	12%

Figure 4.5

Disagreed with a clear reason identified	Total Number	Percentage
X - Better exam prep	11	25%
V - Covers the whole topic	5	11%
W - More warning in advance	5	11%
Y - Grade given	5	11%
Z - Style of questions	5	11%

Figure 4.6

Because there were a larger number of students who agree with the statement in question 7 (Figure 4.4), the data for this gives a more reliable picture of student attitudes. The most common response to give for why the progress check was preferred was that it was less stressful or put less pressure on to the student (Figure 4.5). This response was surprising and had not been a cited by either me or colleagues as a potential benefit to the progress check tasks as opposed to the end of topic tests. Further evidence would need to be gathered to establish why the stress is apparently reduced for progress checks, but a tentative hypothesis would be that this reduced stress is linked to either no grade or percentage being provided

or the naming of the task itself as a progress ‘check’ rather than a ‘test’. The student interviews also provide some evidence for this assertion.

The other reasons tallied more closely with the research literature in terms of improved diagnostic value, an increased coverage of all of the topic studied and its use as a revision tool.

There were fewer clear reasons provided for disagreement to this question and there were many responses that were unique or only give by two students. The most common response was that end of topic tests provided better exam preparation. Interestingly, 11% of students who preferred the end of topic test did so because of its apparent coverage of the whole of a topic, however, this is not something that the end of topic test serves to do and indeed was a factor in changing the assessment from end of topic tests to progress checks. These responses add further confusion to this survey item and again provide evidence that students are confused between the two tasks.

Survey Questions 9 and 10
Q: What do you most like about progress checks? Think about your learning.

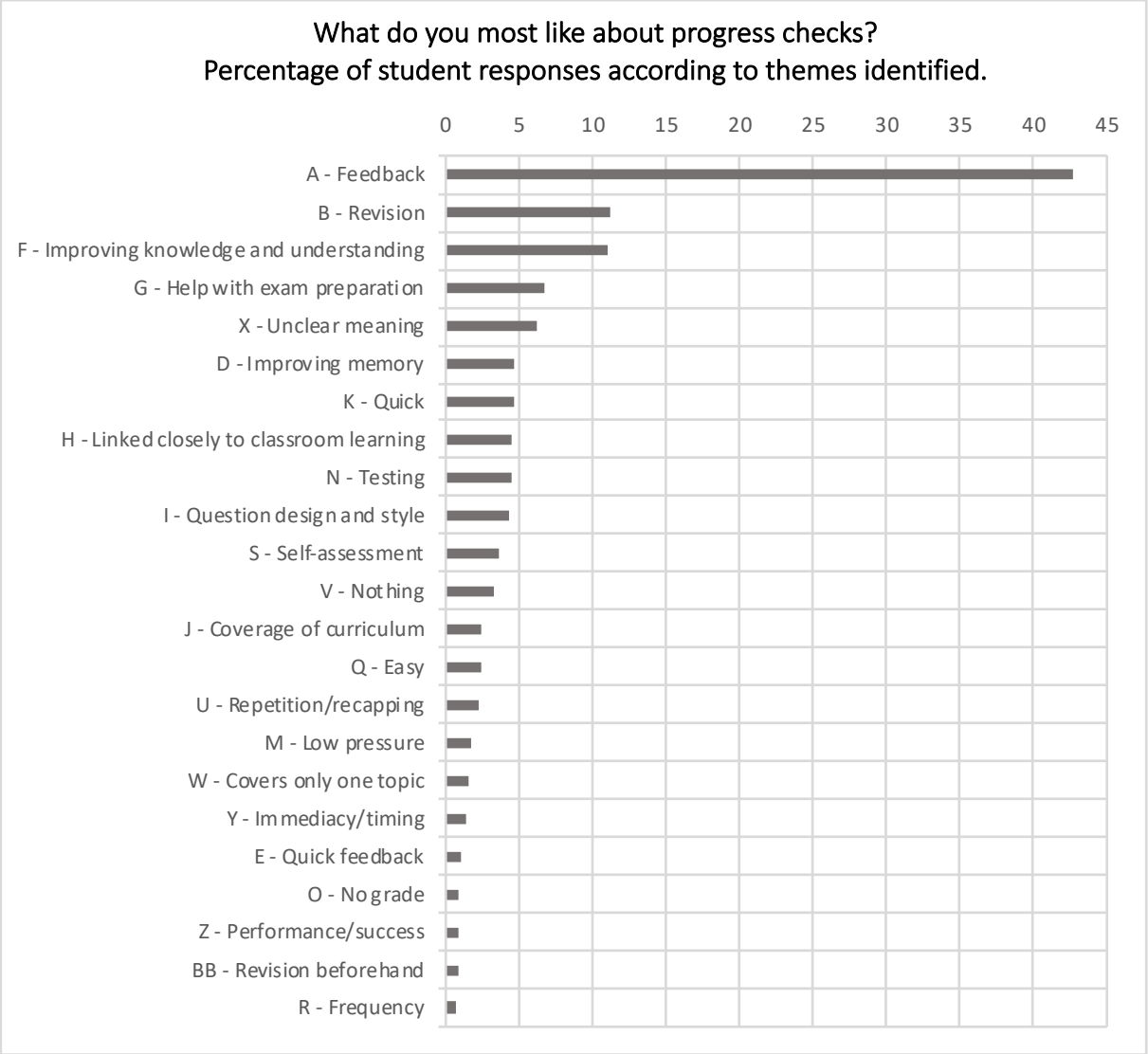


Figure 4.7

By far the most popular response from students to this question was one that focused on the benefits of feedback provided by doing the progress check tasks. 43% of students included this response, a figure much larger than any other factor in both this question and the next. As Figure 4.7 shows, there were a large variety of responses to this question. Perhaps the most surprising result is that only 2% of students identified the less stressful nature of progress checks as being their favourite feature of them, this is in contrast to the 28% who identified this as the reason for preferring progress checks to end-of-topic tests.

Q: What do you not like about progress checks? Think about your learning.

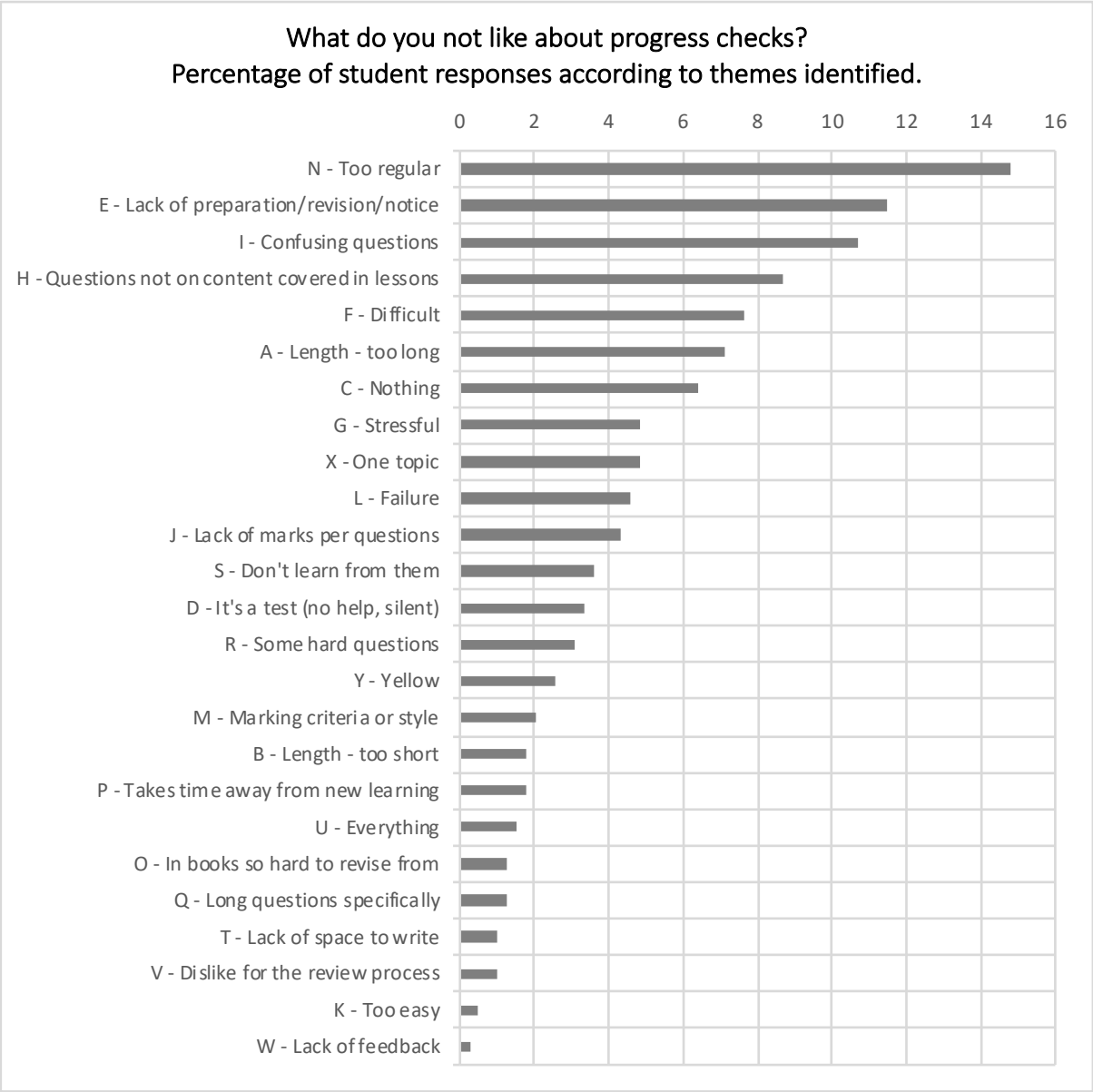


Figure 4.8

This question also prompted a wide variety of response from students. Unlike the previous question, there is no clear consensus provided from students about their least favourite aspects of the intervention. Figure 4.8 shows three themes that were identified by at least 10% of students: the progress checks were too regular; there was a lack of preparation or notice for them; and the questions were confusing.

4.2 Student interviews

The interviews generated themes that were created inside the following categories:

1. **The perceived effect on student learning**

   This is central to the second research question for the project and interviews provide an opportunity to construct a deeper understanding of the impact on students' learning that supplements views generated by the wider student survey.

2. **A comparison between progress checks and end-of-topic tests**

   The students who were interviewed were all in Year 11 and certainly experienced enough in the school to remember end-of-topic tests in order to be able to make a comparison. By asking students to compare, it allows the impact of the progress check to be more clearly delineated.

The perceived effect on student learning

Two themes clearly arose from the data in this category; the students' opinions that the diagnostic function of the test improve learning, and the consequent benefit to revision that the progress checks provided.

Theme One: Diagnostic function

Progress checks are used by students to identify areas of strength and weakness in a curriculum. All of the students interviewed commented on the function of progress checks as being able to diagnose areas of strength and weakness in the topic studied. Comments particularly focused on the diagnosis of weak areas of learning but student views were positive in that these were developmental points rather than absolute judgements on them. There was a general consensus that doing the progress check was enough it itself for students

to understand their current level of understanding and then the subsequent review and teacher feedback further reinforced this:

'*I can see where I went wrong or where I need to develop, I learn from my mistakes.*'

Student 3

'Until I actually try and answer questions about what I'm learning I'm not one hundred percent sure that I know it.'

Student 2

'*You can see where you are, where your weaknesses in that topic are and you can go over and revise. That really helps you understand your areas of improvement*.'

Student 4

Theme Two: Use for revision

Progress checks also appear to cause specific further study and revision that benefits learning. Some students (four out of six) reported revising prior to doing progress checks, suggesting that they may have an effect on learning by merely being present in the curriculum plan. All students interviewed commented on revising topics that come up after doing a progress check, in order to improve their understanding.

'*It promotes students to review beforehand and afterhand (sic.)'*

Student 1

'I really like how specific they are and how I can use them to revise whereas the old ones, I couldn't use it (sic.) that much for revision'

Student 4

Overall, the students were very positive about the impact that progress checks had on their learning, but it is likely that any form of intermediate testing would have had some similar effects. For example, the generation of information about areas of strengths and weakness would be a likely benefit referenced by most students in a UK secondary school familiar with the language of formative assessment. However, there appear to be some more nuanced benefits of the progress checks that are elicited through comparison with end-of-topic tests.

Comparison to end of topic tests

Theme One: Construct coverage

Students commented on the progress checks providing a more comprehensive coverage of the construct. Four out of six students commented specifically on the detailed nature of the progress checks in covering all of the topic recently studied.

'*A lot of the time when you do a test, or some practice questions, you get some bits of the topic and some there are no question on.* (On doing a progress check) *If there's something you don't know, you know you don't know. With other types of testing you could sometimes get the impression that you're one hundred percent o.k. with something when you're not (or the opposite)'*.

Student 2

*'I think the benefit of them is that they are very specific to the topic, so it tries to cover, like, all the little topics inside a big topic so it really helps you whereas that one (*points to end of topic test example) *are not as specific'.*

Student 4

Two students commented on preferring progress checks to end-of-topic tests because they are less stressful:

'They are called progress <u>checks</u> so it's not like failing or passing it's just to see where you are'

Student 1

'I feel like I have more pressure doing end-of-topic tests'

Student 3

Theme Two: Transfer and test preparation

Two disadvantages were raised. Some students felt that there was less opportunity to practise applying knowledge when doing progress checks instead of end-of-topic tests and one student in particular felt quite strongly that they would prefer to have a mark, although this was countered but some other students specifically stating they preferred no mark.

On application:
'I think the progress check is like learning the content and those end-of-unit tests… applying knowledge [pause]… so I think it would be useful to maybe do them both because you can make sure you know all the little bits and pieces and then you can apply that as you do this (points of end-of-topic test example).'

Student 2

On grading:
'I kind of prefer the mark or the grade compared to the progress check where it's just ticks and target [in relation to curriculum descriptors] *'cause I feel like it will give you a realistic idea of what you could of got.'*

Student 6

'*I feel like it's better not to get any marks because that was you look at which questions you need help on, 'cause, if I was to just get a mark, and it it was, like, good, then I would just feel like "oh, I'm good at that topic I don't need to go back on it.'*

Student 5

4.3 Teacher interviews

The relatively open structure and the detail of the answers provided by my colleagues to the questions resulted in a large amount of data that was challenging to categorise. The research question central to this activity was the perceived impact on learning from a teachers' perspective. After listening to and transcribing the responses I chose to categorise the main themes that arose under positive and negative impacts to learning, highlighting those points that appeared frequently across the interviews. There are more positive impacts discussed than negative which reflects the ratio within the data. Overall, there was consensus among the teachers about the intervention in both its negative and positive aspects in terms of impact on students, but there was some divergence in terms of the overall purpose of the progress checks; some teachers commented primarily on the use for them as teachers whereas others almost solely commented on the use of them for the students. Whilst this observation doesn't directly correspond to the research question in focus, this finding will be

discussed further in the conclusion as it does provide meaningful information to me as the head of faculty as we look to develop the progress checks further in the future.

Positive impacts on learning

The progress checks are an effective diagnostic tool

All teachers interviewed commented upon the usefulness of the intervention for them to ascertain the understanding of students in their classes. Whilst all teachers reported this there is some variation in the reasons that they give for the efficacy of the tasks. Three teachers stated that the question design was the reason for the diagnostic worth of progress checks; it was felt that the open structure of the questions meant that students weren't able to provide the answers unless they knew it and a lack of any multiple-choice questions meant that the risk of a false positive induction from the teacher was very low. As one teacher stated, 'the progress checks show what they actually know, with exam questions you can get marks for shallow understanding. You can't do that with the progress check questions.' Other teachers commented upon the thoroughness of the questions covering all of the construct as being a reason for their increased diagnostic value, particularly compared with the end-of-topic tests.

Positive impacts on the self-regulatory behaviours of students

Some teachers described improved learning behaviours in their students that they attributed directly to the use of the progress checks. In their observation of this behaviour one teacher explained that '*they learn to be reflective learners, they can't wait to start improving on their work'*. Another teacher stated that '*an increased number of students are starting to revise*' explaining that, in making the students realise what they do and don't know it '*increases ownership of learning'*. One teacher hypothesised that the regular, low-stakes testing that the

progress checks provides may help prepare students psychologically for higher-stakes examination in the future: '*the students get used to testing and don't get overwhelmed when they have their end of cycle tests. Progress checks are more demanding due to the nature of the question, so pupils are more confident in end of cycle tests now*'.

An emphasis on review and adapting teaching benefits learning

Part of the intervention saw the curriculum structure of a teaching unit change to give dedicated review time at the end of the unit of work post-progress check. Teachers felt that this review time impacted very positively on the learning of their students. The main reasons for this were that it afforded teachers opportunity to re-teach areas of the topic that the progress check had highlighted as an issue, and secondly that it gave students time to reflect and develop self-assessment skills upon reviewing their own responses and improving them after receiving feedback.

Negative impacts

Lower attaining students have, at times, not coped well with the intervention

Four teachers expressed concerns that lower ability students have struggled to access and benefit from progress checks in the way that higher ability students have. The primary reason for this was the lack of ability in some students to access the questions and provide a meaningful response. Teachers felt that, in the end-of-topic tests that took place before progress checks were introduced, students were able to at least attempt some questions in the form of ticking a box or drawing lines between boxes. However, because the progress checks do not allow for this, there may have been a negative impact on their motivation. As one teacher stated, 'easier questions aren't better diagnostically but they might increase

motivation overall' so there is a fine balance that needs to be struck in the future development of the tasks.

4.4 Summative data analysis

Key Stage 3 (Years 7 and 8)

Exam Pairing	Mean exam score	Difference in mean	t	df	Sig. (2-tailed)
Y8 Biology Foundation 2017	19.8	-1.0	0.93	71	0.356
Y8 Biology Foundation 2018	18.8				
Y8 Chemistry Foundation 2017	24.9	-2.7	2.637	70	0.01***
Y8 Chemistry Foundation 2018	22.2				
Y8 Physics Foundation 2017	13.3	-0.7	0.858	71	0.394
Y8 Physics Foundation 2018	12.6				
Y8 Biology Higher 2017	21.8	-2.1	2.356	79	0.021**
Y8 Biology Higher 2018	19.7				
Y8 Chemistry Higher 2017	23.1	-2.0	2.563	77	0.012**
Y8 Chemistry Higher 2018	21.1				
Y8 Physics Higher 2017	19.1	+1.8	-2.118	78	0.037**
Y8 Physics Higher 2018	20.9				
Y7 Biology Foundation 2017	18.1	+1.8	-1.859	78	0.067*
Y7 Biology Foundation 2018	19.9				
Y7 Chemistry Foundation 2017	17.5	+0.4	-0.469	78	0.64
Y7 Chemistry Foundation 2018	17.9				
Y7 Physics Foundation 2017	12.0	-2.1	2.436	76	0.017**
Y7 Physics Foundation 2018	9.9				
Y7 Biology Higher 2017	21.7	-2.4	1.726	82	0.088*
Y7 Biology Higher 2018	19.3				
Y7 Chemistry Higher 2017	21.9	-2.3	2.752	81	0.007***
Y7 Chemistry Higher 2018	19.6				
Y7 Physics Higher 2017	14.9	-1.2	1.423	81	0.159
Y7 Physics Higher 2018	13.7				

Figure 4.9

* significant at 10% level ** significant at 5% level *** significant at 1% level

Comparison of attainment in end-of-year examinations showed an improvement in two exams but a worsening of performance in six exams (Figure 4.9) for students in Year 7 and Year 8. Those results with large T-values (either negative or positive) and less than a 5% significance indicate the strong possibility of a real change in performance with those significant at 10% notable but not conclusive (Cohen et. al, 2018).

The data provides mixed evidence but overall suggests that students in years 7 and 8 performed less well than the previous year group did on this set of exams. It is very difficult

to attribute any one cause to this finding; whilst the examinations that the students took remained the same across the two years, different teachers, different setting arrangements and different prior attainments of each year group are all factors that may have contributed alongside the change in medium-cycle formative assessment practice. However, the prior attainment of these particular Year 7 and 8 groups with their respective peer groups in previous years is very closely aligned (Appendix G) and, whilst there were changes in teachers from one year to the next, in the Year 7 teaching team there was only one change in teacher and in Year 8 there were two. In the previous teaching year students were set by ability but during this year they were in mixed ability groupings which is another factor to consider but one that is beyond the scope of this research. Despite these other variables existing, there is reason to think that they were as closely controlled as is possible in a school, so this evidence could be used to tentatively suggest that the intervention had a negative impact on end-of-year examination attainment, this is particularly the case for the results significant at 5% and below.

Key Stage 4 (Years 9 and 10)

Exam Pairing	Mean exam score	Difference in mean	t	df	Sig. (2-tailed)
Y9 Biology Foundation 2017	12.8	+1.8	-2.533	74	0.013**
Y9 Biology Foundation 2018	14.6				
Y9 Chemistry Foundation 2017	16.1	+2.7	-2.958	75	0.004***
Y9 Chemistry Foundation 2018	18.8				
Y9 Physics Foundation 2017	8.1	+2.5	-3.022	68	0.004***
Y9 Physics Foundation 2018	10.6				
Y9 Biology Higher 2017	15.4	+0.7	-0.526	53	0.601
Y9 Biology Higher 2018	16.1				
Y9 Chemistry Higher 2017	19.7	+1.8	-1.311	53	0.195
Y9 Chemistry Higher 2018	21.5				
Y9 Physics Higher 2017	14.7	+1.3	-0.918	49	0.363
Y9 Physics Higher 2018	16.0				
Y10 Biology Foundation 2017	10.6	+1.1	-1.653	49	0.105
Y10 Biology Foundation 2018	11.7				
Y10 Chemistry Foundation 2017	10.1	-0.3	0.548	47	0.586
Y10 Chemistry Foundation 2018	9.8				
Y10 Physics Foundation 2017	8.7	+1.9	-1.959	46	0.056*
Y10 Physics Foundation 2018	10.6				
Y10 Biology Higher 2017	20.9	+2.6	-1.984	47	0.053*
Y10 Biology Higher 2018	23.5				
Y10 Chemistry Higher 2017	20.0	+5.8	-3.102	45	0.003***
Y10 Chemistry Higher 2018	25.8				
Y10 Physics Higher 2017	13.0	+11.9	-8.944	47	0.000***
Y10 Physics Higher 2018	24.9				

Figure 4.10

* significant at 10% level ** significant at 5% level *** significant at 1% level

The data from examinations sat by Year 9 and 10 students provides a contrasting picture from that of Year 7 and 8. Figure 4.10 indicates an almost exclusively positive change in attainment, with the only decrease in mean score not proving to be statistically significant (Year 10 Chemistry Foundation paper). Smaller differential factors for these data are caused by lower numbers in these particular year groups compared with Year 7 and 8. In Year 10 there are particularly large increases in mean score which result in significant improvements.

The extraneous variables identified previously (setting, teacher changes, and prior attainment) must also be considered for this data set. Unlike Year 7 and 8, the way in which students were grouped remained the same for this cohort of students but there were greater changes in the teaching teams and prior attainment is more variable across these two cohorts in comparison to the students previously in these year groups. In Year 10 particularly, the teaching team was quite different to that of the previous year (both in terms of personnel and also which groups were taught by individuals) so individual teacher effectiveness was not well controlled. The prior attainment data for Year 10 for this academic year is quite different to that of the groups below (Year 9) and above them (Year 11) (Appendix G) so it would be expected for this year group to out-perform the past cohort on these examinations. However, by this reasoning it could also be argued that the data for Year 9 would show a decrease in performance because of the much higher prior attainment of Year 10, but this is not the case. Particularly for the foundation students in Year 9, they have appeared to of out-performed the previous year when they would be expected to achieve less well than them.

4.5 Discussion

Research question one: What impact does the assessment intervention, designed to improve learning over time, have on summative grades for students in years 7, 8, 9 and 10?

The literature on formative assessment suggests that effective implementation will result in improved learning over time (Black and Wiliam, 1998). However, it has also been noted that measurable improvements in learning are very under-represented in the literature (Bennett, 2011) and much research does not contain a quantitative component of this nature (Baird et al., 2014). This provides good reason for using this research question that attempts to quantify the effects of the intervention so that the opinions offered by students and teachers may be supported or challenged by it.

The exams that students sat at the end of the year were entirely separate from the progress check intervention process except for covering the same curriculum. The data suggest that there is a difference in impact for students in Years 7 and 8 compared with those in Years 9 and 10. Figure 4.9 shows that the exam performance for Year 7 and 8 students is less good compared with the previous year. In contrast, Year 9 and 10 students appear to have done considerably better than their peers. The survey data does not give the same conclusions; students in Years 7 and 8 mostly felt confident that the progress checks were benefitting their exam performance. 67% of students in these year groups reported to either agree or strongly agree with the statement 'progress checks help me to do better in exams' compared with 57% of students in Years 9 and 10, who actually appear to have done significantly better.

Popham (2003) explains that the purpose of any test is to generate evidence, by a 'students' overt test performance' that can be used by teachers to make inferences about a 'student's covert status' (p.4). It is possible that the exams sat by the two groups of students (Years 7

and 8; Years 9 and 10) varied in their ability to generate this view of the covert status described by Popham. The exams sat by the two groups were designed in quite different ways. Whilst the Year 9 and 10 exams used sample papers provided by the relevant exam board, the Year 7 and 8 papers were made in the faculty by various members of staff and from a variety of sources. This raises issues of validity for the data used to make conclusions to the research question.

The validity of inferences from the Year 7 and 8 exams should be questioned more than the Year 9 and 10 exams. Because they were made 'in-house' rather than as full mock papers that will have undergone pre-testing after having been written by professional exam-writers, they have not undergone the same level of scrutiny. One factor that was not systematically considered in the making of the Year 7 and 8 exams was the difficulty of each question which meant that some exams contained a disproportionate amount of either very difficult or very easy questions. Another issue that has arisen from these exams is that some of the questions are not well-aligned to the curriculum that we teach. Some questions, for example, are taken from old SAT papers that were designed for the previous key stage three curriculum which differs from the one followed now.

Despite these potential issues with the examinations that the students sat, the literature would suggest that there should still have been an observable improvement in the learning of students over the academic year. Low-stakes assessment in the classroom can potentially provide two clear benefits to learning. The first is that the process of retrieval through testing improves learning (Einstein, Mullet and Harrison, 2012; Karpicke, Butler and Roediger, 2008; McDaniel et al., 2011; McDaniel et al., 2013; Roediger and Butler, 2011), and the second is that accurate identification of individual strength and weakness provides feedback that both

teachers and students can use to improve. The literature suggests that the improvements in learning that results through these two mechanisms would then result in a better performance in an exam designed to judge how successfully a student has learnt over the school year.

As discussed previously, there were extraneous variables that could not have been controlled in a normal educational environment which may have caused the differences in achievement for the two groups of students. However, these variances may be said to affect both sets of data. There is a more obvious difference between these two cohorts that set them apart; the age of the students in them. This was not a factor not explored in the literature and further investigation shows this to be an area where little work is published in relation to formative assessment, and so provides scope for future research.

Research question two: What are students' views of the impact of the assessment intervention on their learning over time?

Despite the conflicting evidence from exam data, other sources of evidence gathered suggest that the intervention, according to students' views, have had a largely positive impact on learning across both key stage three and four and across ability ranges. The reasons for this improvement to learning fit well alongside the literature in terms of what successful formative assessment looks like in practice.

The literature suggests that effective formative assessment should provide a direction for learning, both prior to the assessment and after it (Brookhart, 2018; Dann, 2014; Wiliam, 2011b). As Baird et al. (2014a) describe, assessment 'wash back' means that the assessment processes that students go through will dictate the focus of the teaching and learning that come before it. For this reason, it is important that there is knowledge and trust, from both teachers and students that the assessment tasks used to measure learning are closely linked with the learning that has taken place beforehand. There are roles for both teachers and learners in clarifying the direction of learning (Wiliam 2011a) and student views collected from both surveys and interviews suggest that this direction has been successfully shared. 87% of students agreed or strongly agreed that they understand why they do progress checks in science which suggests that teachers across the faculty have effectively communicated the purpose of the intervention, a key component in successfully establishing routines and processes in the classroom that students understand and see worth in. Cizek (2010) identifies a criterion of 'clear, specific learning goals' as an indicator of successful formative assessment and the evidence supports that this has been achieved.

The evidence from student interviews and surveys also indicates a widely-held view that the intervention has had a positive impact on learning with 77% of students think that doing progress check tasks have helped them learn. In interviews students identified the diagnostic function of the tasks as being the most helpful to their learning and this is substantiated by 43% of students across the school stating this as the feature that they most liked about doing a progress check. Alongside many students identifying it as a helpful feature, 79% of students agreed that they know how to improve following the assessment task; this involvement by students in the assessment process was identified as an under-developed aspect of formative assessment in previous research in the school (Author, 2017). Feedback that both students and teachers can use to improve learning is identified within the literature as a key component of effective formative assessment (Wiliam, 2011a, p. 45) and this evidence suggests that the task design has been effective (Wiliam, 2014, p.8) in enabling this decision-making process.

In student interviews, Year 11 students commented upon the quality of the feedback from the progress checks in reinforcing the learnt domain and their responses demonstrated a trust in the process where they are confident that everything they need learn is included in the progress check. This suggests that the assessment is not only useful in clarifying the curriculum aims (Popham, 2003) but a trust in assessment processes is likely to improve the self-regulatory behaviour of students so that they undertake action to improve their learning as a result of feedback. This was reported by Year 11 students in interviews and teachers also report seeing this behaviour developing in students across the school, but student surveys suggest that this is not yet well-developed across younger year groups.

When interviewed, some students were able to articulate changes to the validity of judgements made from these interim tests compared with previous end-of-topic tests. In explaining 'If there's something you don't know, you know you don't know' Student 2 speaks to the tasks achieving a level of the three validity measures proposed by Pellegrino, DiBello and Goldman (2016).

However, this claim is only partially evidenced by the data. The cognitive validity of judgements may be more problematic for lower-ability students or those in earlier year groups in the school. 11% of students described question difficulty as being the feature they least liked which suggests that some questions may have acted as a cognitive barrier (perhaps through inaccessible language or ambiguity) which then prevented the student or teacher to make inferences of high validity. This view is supported by four teachers in their interviews who expressed concern that lower ability students found the tasks very difficult at times. However, the exam attainment data clouds this point further by suggesting that students taking foundation papers have not done less well than their peers sitting higher papers. In the case of Year 9, the foundation students actually achieved a greater improvement in test results than those sitting the higher paper. This may suggest that, whilst the cognitive validity of inferences is reduced because of the style of questioning, there is still benefit to the student in doing the progress check. This is supported by research on the testing effect as improving long-term retention through the practice of retrieval (Einstein, Mullet and Harrison, 2012; Karpicke, Butler and Roediger, 2008; McDaniel et al., 2011; McDaniel et al., 2013; Roediger and Butler, 2011).

Students widely reported the decreased levels of stress that they experience doing progress checks in comparison to previous end-of-topic tests. This was a surprising result and the

reasons for it are not clear although one student interviewed suggested that the name of the intervention was a reason for this alongside the non-judgmental marking system.

9% of students identified questions that asked about content they had not studied as a part of progress checks that they did not like. Whilst it is of course possible that the students were unable to identify this correctly, there is some evidence to suggest that it may be true as one teacher reported this to be a problem in their interview. This highlights the important of using assessment as a bridge between what should be taught in order to be learnt, identified by Wiliam as 'clarifying, sharing and understanding learning intentions and success criteria' (2011a, p. 46).

The progress check tasks are designed to be reusable resources that students can use to help themselves self-quiz when revising. The literature suggests that this low-stakes repeated retrieval can have long-term benefits for both definitional and transfer test items (McDaniel, 2013). All students who were interviewed reported on progress checks being useful for this however only 40.5% of students in years 7-10 agreed that they had used the tasks for revision. This highlights a difference in use that may warrant further exploration; but it would be reasonable to assume that a high ability groups of students close to taking their GCSE examinations would report higher levels of revision through all techniques than their peers lower down in the school. The importance of 'activating students as owners of their own learning' (Wiliam, 2011a, p.46) is perhaps an integral part of the process that needs more attention in the future.

The student interviews and surveys also raised evidence that students are worried about how well prepared they will be for exams, particularly in relation to transferring and applying their

knowledge to different contexts. The literature would suggest that students shouldn't fear this as much as they might (Dunlosky, et al. 2013) but it is an aspect of the intervention that will need to be monitored in the future.

Research question three: What are teachers' views of the impact of the assessment intervention on their students' learning over time?

Overall, teachers were very positive about the potential impacts that the progress checks might have on student learning over time and were generally positive that this potential had been achieved across the school year.

Like students, the diagnostic capabilities of the tasks were identified by teachers as a key feature in improving learning. There were three elements to the feedback that the teachers reported: feedback for the students on their progress; feedback to the teacher about their students' progress; and feedback to the teacher on the efficacy of their own teaching. In addition, teachers also commented on the improved inferential validity (Pellegrino, DiBello and Goldman, 2016) that was enabled by the tasks. This highlights two elements of the process that Black and Wiliam (2012) claimed was important in the effective implementation of formative assessment; the use of well-designed instruments for the process and establishing students as part of the process.

Previous research at School X (Author, 2017) showed that there was a focus on teacher information and teacher action when teachers were asked about the components of formative assessment. This viewpoint remained evident; for two teachers, the main impact of the progress check tasks was to improve their understanding of their students' needs and then acting in order to improve them. Whilst this is a positive outcome, it must be considered alongside the literature that stresses the importance of students becoming part of the assessment process (Wiliam, 2011a), developing self- and peer-assessment skills and learning behaviours that enable them to think about their own thought processes to become self-regulating (Zimmerman, 2000).

Bennett (2011) outlined issues with the implementation of formative assessment in schools. The definitional issue described confusion between formative assessment as either instrument or process. The responses from teachers highlighted the effectiveness of progress checks as an instrument that allowed them opportunity to then improve learning. Some teachers doubted their ability to improve the learning of their students following the progress check task. This may be compared with Bennett's professional development issue (2011) and indicates that a focus on the review lessons that follow the progress check may be required in future faculty discussions (Heritage, 2018).

Teachers frequently spoke of the ability of the task to be used later in revision, both to identify areas of development but also as practice questions for retrieval practice. Teachers reported students using the progress checks more frequently for revision when compared with the previous end-of-topic tests and also reported higher levels of engagement from students with the assessment process. This suggests that the progress check encourage self-regulation behaviours in students (Zimmerman, 2012) and the nature of the tasks mean that they encourage revision behaviours supported by research (Dunlosky et al., 2013).

Chapter 5

Conclusions and Implications

5.1 Conclusions

This research and development project aimed to combine research on formative assessment with that of cognitive science in order to modify, and hopefully improve the formative assessment practices of the science faculty at School X.

Previous research suggests important roles for teachers and students in successfully implementing formative assessment strategies (Black and Wiliam, 1998) and the processes that are carried out by these groups can be described as establishing 'where the learner is going; where the learner is now; and how to get there' (Wiliam, 2011a, p. 46). The research project sought to improve a central tool in this process, particularly in creating medium-cycle assessment opportunities that 'elicit evidence of learning' (Wiliam, 2011a, p.46) that not only improve the quality of evidence gathered for both students and teachers, but also served as a learning opportunity in itself (Shepard, Penuel and Pellegrino, 2018b). It was hoped that the focus of our medium-cycle assessment tasks could be shifted in the way that Shepard, Penuel and Pellegrino describe:

> 'Instead of "scores," which offer teachers little information about what to do next, it is much more important that formative assessment questions, tasks, and activities provide instructional "insights" about student thinking and about what productive next steps might be taken.'
>
> (Shepard, Penuel and Pellegrino, 2018b, p. 27)

The research questions focused on the impact to learning that the intervention had, using the views of teachers and students alongside exam data to make this analysis. In the analysis of the evidence it is clear that, for both students and teachers, the progress checks have been

successful in their diagnosis of strengths and weaknesses in individual students. In teacher interviews, student interviews and student surveys, the diagnostic ability of the progress checks was frequently identified as effective, particularly in relation to the tasks that were used prior to the intervention. There is also evidence to suggest that the inferences from these tasks are now more valid (Pellegrino, DiBello and Goldman, 2016) and thus more useful, than those made from the previous tool used for the process. The improved inferential validity is achieved through improving the cognitive and instructional validity of judgements made. The data from teacher and student interviews suggest that the biggest improvements have come in improving the instructional validity of the assessment tasks which has been achieved in the following ways: the tasks covering all of the taught construct rather than a sample; a reduced variation in the difficulty of questions asked and the nature of the questions as open so that the answers cannot be correctly guessed. Student surveys and teacher interviews suggest that cognitive validity could be improved further by making some questions more accessible in their wording for students with lower prior attainment.

Further to improved diagnosis, students appear to engage with the formative process as a whole and report an understanding of the process of which the progress checks are a part. Developing self-regulating students and involving them with the assessment process is key to the success of the intervention (Zimmerman and Moylan, 2009) and the evidence suggests that, for older students close to exams, the progress checks have improved their self-regulation as they use the tasks for revision purposes. However, the evidence for this in younger students is weak, which provides opportunity for future development.

It appears that students are more comfortable with this format of assessment than they were with previous tests, with many reporting lower levels of stress associated with progress

checks. One teacher also reported observing students being more confident in summative examination, which they attributed to the effect of the progress check in building resilience and confidence.

Some students expressed concerns that the intervention reduced the amount of preparation they had for summative examination. In interviews, students were concerned that the lack of application and transfer questions on the progress check tasks meant that they had less practice in this area. The research evidence suggests that this may not be as damaging to learning as the students may think (Dunlosky et al., 2013), but it is important to consider where these opportunities may come elsewhere in the curriculum or whether this is a symptom of the high-stakes accountability culture that both students and teachers are deeply entrenched in.

It is very difficult to attribute the changes to exam performance to a cause and the data that suggests slightly worse exam performance for Years 7 and 8 and significantly better performance for Years 9 and 10 should be treated with scepticism. The increase in exam performance in the older age group is supported by previous research into the effects of low-stakes retrieval practice (Einstein, Mullet and Harrison, 2012; Karpicke, Butler and Roediger, 2008; Littrell-Baez et al., 2015), and the literature suggests that this effect may be compounded further by making this a part of formative assessment that encourages reflection on learning at all stages. Conversely, the evidence from the younger age group directly contradicts this. The extraneous factors that are inevitable across a school faculty and across different school years also render these findings tentative at best.

It is one thing to successfully feedback to a student or teacher about what has and hasn't been learned successfully but the actions that students and teachers must take following this diagnosis may be more challenging still. Research by Heritage et al. (2009) found that 'teachers are better at drawing reasonable inferences about student levels of understanding from assessment information than they are at deciding the next instructional steps' (p. 24). There is some evidence to suggest that this process should be the next focus of the faculty in the development of formative assessment practices. The summative assessment data shows that there may be inconsistency in actually improving understanding in students following assessment so that the potential improvement in learning is not realised by reflective action by either the student or teacher. An alternative explanation is there was a difference in the ability of the summative examinations to measure the learning of students.

The design of the progress checks was heavily influenced by learning science research but the features that were identified by teachers and students as contributing most to learning rarely attributed improvements to these factors. This may be because of the implicit nature in which the research was applied to the assessment design. For example, teachers and students may not be aware of the testing effect so are not in a position to attribute cause to it. Dunlosky et al. (2013) would suggest that a wider education of effective study techniques that include retrieval practice may not only allow teachers and students to better appreciate the effects but may also encourage the study body to employ them more frequently.

5.2 Implications for future practice and research

The intervention changed a central practice of a faculty and it has been decided that, because of the largely positive outcomes of the intervention, we will continue to use progress checks as a tool in medium-cycle formative assessment for the foreseeable future. This is pleasing because a large amount of work has gone into developing the tasks themselves, ensuring consistent implementation across eight different teachers and developing the skills of less experienced staff in interpreting evidence in order to improve future instruction.

In response to the evidence gathered, progress checks for lower-attaining students will be developed next year to provide more structure whilst attempting to not compromise on inferential validity. This would also seek to improve the cognitive validity of assessments for these students (Pellegrino, DiBello and Goldman, 2016). Whilst there was little quantitative evidence to prove a disadvantage for lower-attaining students, teachers felt strongly that the motivation of these students may have been reduced by the intervention and it may be beneficial to sacrifice an element of validity in order to improve motivation and then, possibly, self-regulation in students (Zimmerman and Moylan, 2009). The contribution of sociocultural theories of learning also remind us that it is important to consider this aspect of learning.

There is some evidence to suggest that improvements still need to be made to the way that teachers and students are making improvements to teaching and understanding following the progress checks. Research indicates that this is a more difficult aspect of formative assessment to achieve (Heritage, 2009). The work of the faculty in the next academic year will be focused on modifying and establishing best practice for the review lessons so that feedback from the progress check leads to better improvement of learning.

Interest from other faculty leaders in the school also gives an opportunity to see if the principles used within science can transfer to other subjects well. The research emphasises the need for domain-specific approaches (Bennett, 2011) with generic approaches often failing to produce the expected improvements so this will be approached with caution. Further research may also be undertaken as a result of the difference in summative exam results between years 7 and 8 and years 9 and 10. The results suggest that the intervention has been much more useful for the older age group and would provide interesting scope for another research project.

Reference List

Agarwal, P.K., Bain, P.M. & Chamberlain, R.W. (2012) The Value of Applied Research: Retrieval Practice Improves Classroom Learning and Recommendations from a Teacher, a Principal, and a Scientist. *Educational Psychology Review*, 24 (3), 437–448.

Alderson, J. C. and Wall, D. (1993) Does washback exist? *Applied Linguistics*, 14, 115–129.

Andersson, C. and Palm, T. (2017) Characteristics of improved formative assessment practice. *Education Inquiry*, 8 (2), 104–122.

Author (name withheld) (2017) Analysing teachers' perspectives of formative assessment. Unpublished.

Baas, D. et al. (2015) The relation between assessment for learning and elementary students' cognitive and metacognitive strategy use. *British Journal of Educational Psychology*, 85 (1), 33–46.

Baddeley, A. D., Eysenck, M. W., and Anderson, M. C. (2009) *Memory*. Hove: Psychology Press.

Baird, J. A. et al., (2014a) State of the Field Review Assessment and Learning. *Report for the Norwegian Knowledge Centre for Education*, Oslo: Knowledge Centre for Education.

Baird, J. A. et al., (2014b) Teachers' views on assessment practices. *Assessment in Education: Principles, Policy and Practice*, 21 (4), 361–364.

Baird, J. A. et al., (2017) Assessment and learning: fields apart? *Assessment in Education: Principles, Policy and Practice*, 24 (3), 317–350.

Barton, G. (2018) Ofsted doesn't sack headteachers – but it is part of an unforgiving accountability system. [Online]. *Times Educational Supplement,* 8 February. Available from https://www.tes.com/news/ofsted-doesnt-sack-headteachers-it-part-unforgiving-accountability-system [Accessed 29 August 2018]

Bell, B. and Cowie, B. (2001) *Formative assessment and science education*. Dordrecht, The Netherlands: Kluwer Academic Press.

Bennett, R. E. (2010) Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research & Perspective*, 8 (2–3), 70–91.

Bennett, R.E. (2011) Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18 (1), 5–25.

Bennett, R.E., (2015) The Changing Nature of Educational Assessment. *Review of Research in Education*, 39 (1), 370–407.

Birenbaum, M. et al., (2015) International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, 13 (1), 117–140.

Bjork, R. A. and Bjork, E. L. (1992) A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn and R. Shiffrin (eds.) *From learning processes to cognitive processes: Essays in honor of William K. Estes*, Hillsdale, NJ: Erlbaum. 35-67.

Bjork, R. A., Dunlosky, J. and Kornell, N. (2013) Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology*, 64 (1), 417–444.

Black, P. et al. (2003) *Assessment for Learning - putting it into practice.* Maidenhead, U.K.: Open University Press.

Black, P. et al. (2004) Working inside the Black Box: Assessment for Learning in the Classroom. *Phi Delta Kappan*. 86 (1) 8-21.

Black, P. (2015) Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy and Practice*, 22 (February), 37–41.

Black, P. and Wiliam, D. (1998) Inside the Black Box: raising standards through classroom assessment. *Phi Delta Kappan*. 80 (2) 139-144.

Black, P. and Wiliam, D. (2012) Developing a theory of formative assessment. In: Gardner. J (ed.), second edition, *Assessment and Learning Educational Assessment,* London: Sage. 206-229.

Blakemore, S. J. and Bunge, S.A. (2012) At the nexus of neuroscience and education. *Developmental Cognitive Neuroscience*, 2 (1), 1–5.

Blakemore, S. J (2018) Editor note. *Impact – Journal of the chartered college of teaching.* 2. 1-3.

Box, C., Skoog, G. and Dabbs, J.M. (2015) A Case Study of Teacher Personal Practice Assessment Theories and Complexities of Implementing Formative Assessment. *American Educational Research Journal*, 52 (5), 956–983.

British Educational Research Association [BERA] (2018) *Ethical Guidelines for Educational Research*, fourth edition, London. Available from: https://www.bera.ac.uk/researchers-resources/publications/ethical- guidelines-for-educational-research-2018 [Accessed 29 August 2018]

Brookhart, S.M. (2018) Learning Is the Primary Source of Coherence in Assessment. *Educational Measurement: Issues and Practice*, 37 (1), 35–38.

Brown, G. T. L. (2004) Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11 (3), 301–318.

Bruer, J. T. (1997) Education and the brain: A bridge too far. *Educational Researcher*, 26 (8), 4–16.

Bruner, J. S. (1977) *The process of education*. Cambridge: Harvard University Press.

Christodoulou, D. (2014) *Seven myths about education*. Oxon: Routledge.

Christodoulou, D. (2016) *Making good progress? The future of Assessment for Learning*. Oxford: Oxford University Press.

Clark, I. (2012) Formative Assessment: Assessment Is for Self-regulated Learning. *Educational Psychology Review*, 24 (2), 205–249.

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

Coffey, J. E. et al., (2011) The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching*, 48 (10), 1109–1136.

Cohen, L., Manion, L., and Morrison, K. (2018) *Research methods in education.* 8th ed., New York: Routledge.

Creswell, J. W. and Plano Clark, V. L. (2011) *Designing and conducting mixed methods research.* 2nd ed., Thousand Oaks, CA: Sage.

Crisp, G.T. (2012) Integrative assessment: Reframing assessment practice for current and future learning. *Assessment & Evaluation in Higher Education*, 37 (1), 33–43.

Crooks, T. (2011) Assessment for learning in the accountability era: New Zealand. *Studies in Educational Evaluation*, 37 (1), 71–77.

Dann, R. (2014) Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, 21 (2), 149–166.

Denscombe, M. (2017) *The good research guide: For small-scale social research projects*. 6th ed., London: Open University Press.

Dubinsky, J. M., Roehrig, G. and Varma, S. (2017) Infusing neuroscience into teacher professional development. *Educational Researcher*, 42 (6), 317–329.

Dunlosky, J. et al. (2013) Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, Supplement*, 14 (1), 4–58.

Education Endowment Foundation (2018) *Embedding Formative Assessment*. [Online]. Available from: https://educationendowmentfoundation.org.uk/pdf/generate/?u=https://educationendowmentfoundation.org.uk/pdf/project/?id=252&t=EEF%20Projects&e=252&s= [Accessed 21 August 2018]

Einstein, G. O., Mullet, H. G. and Harrison, T.L. (2012) The Testing Effect: Illustrating a Fundamental Concept and Changing Study Strategies. *Teaching of Psychology*, 39 (3), 190–193.

Elliot, V. et al. (2016) *A marked improvement? A review of the evidence on written marking*. London: Education Endowment Foundation, University of Oxford.

Gardner, J. et al. (2010) *Developing teaching assessment.* Berkshire, England: Open University Press.

Great Britain, Department for Children, Schools and Families (2008) *The assessment for learning strategy*. [Online]. Available from: http://dera.ioe.ac.uk/8161/7/DCSF-00341-2008_Redacted.pdf [Accessed 5 September 2018]

Harlen, W. and James, M. J. (1997) Assessment and Learning: differences and relationships between formative and summative assessment. *Assessment in Education,* 4 (3), 365-380.

Hartley, J and Betts, L. R. (2010) Four layouts and a finding: the effects of changes in the order of the verbal labels and numerical labels on Likert-type scales. *International Journal of Social Research Methodology,* 13 (1), 17-27.

Hattie, J. and Timperley, H. (2007) The Power of Feedback. *Review of Educational Research*, 77 (1), 81–112.

Hayward, L. (2015) Assessment is learning: the preposition vanishes. *Assessment in Education: Principles, Policy and Practice*, 22 (1), 27–43.

Heritage, M. et al. (2009) From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice*, 28 (3), 24–31.

Heritage, M. (2018) Making Assessment Work for Teachers. *Educational Measurement: Issues and Practice*, 37 (1), 39–41.

Independent Teacher Workload Review Group (2016) *Eliminating unnecessary workload around marking.* [Online]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/511256/Eliminating-unnecessary-workload-around-marking.pdf [Accessed 29 August 2018]

James, M. and Gipps, C. (1998) Broadening the basis of assessment to prevent the narrowing of learning. *The Curriculum Journal*, 9 (3), 285–297.

James, M. (2006) Assessment, Teaching and Theories of Learning. In: Gardner, J. (ed) *Assessment and Learning.* London: Sage. 47-60.

Johnson, R., Onwuegbuzie, A. J. and Turner, L. A. (2007) Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1 (2), 112-133

Karpicke, J. D. and Roediger, H. L. (2008) The critical importance of retrieval for learning. *Science*, 966, 13–15.

Karpicke, J. D., Butler, A. C. and Roediger, H. L. (2009) Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, 17 (4), 471–479.

Karpicke, J. D. and Blunt, J. R. (2011) Retrieval practice produces more learning than concept mapping. *Science*, 331 (6018), 772–775.

Kingston, N. and Nash, B. (2011) Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30, 28–37.

Kirschner, P. A., Sweller, J. and Clark, R. E. (2006) Why minimal guidance during instruction does not work. *Educational Psychologist*, 41 (2), 87–98.

Koretz, D. (2008) *Measuring Up*. London: Harvard University Press.

Leahy, S. et al. (2005) Classroom assessment: Minute by minute, day by day. *Educational Leadership*, 63 (3), 18–24.

Littrell-Baez, M. K. et al. (2015) Using retrieval practice and metacognitive skills to improve content learning. *Journal of Adolescent and Adult Literacy*, 58 (8), 682–689.

Liu, Y. et al. (2017) Differentiation of subsequent memory effects between retrieval practice and elaborative study. *Biological Psychology*, 127 (23), 134–147.

McDaniel, M. A., Roediger, H. L. and McDermott, K. B. (2007) Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin and Review*, 14 (2), 200–206.

McDaniel, M. A. et al. (2011) Test-Enhanced Learning in a Middle School Science Classroom: The Effects of Quiz Frequency and Placement. *Journal of Educational Psychology*, 103 (2), 399–414.

McDaniel, M.A. et al. (2013) Quizzing in Middle-School Science: Successful Transfer Performance on Classroom Exams. *Applied Cognitive Psychology*, 27 (3), 360–372.

Nichols, P. D., Meyers, J. L. and Burling, K. S. (2009) A framework for evaluating and planning assessments intended to improve student achievement. *Educational Measurement: Issues and Practice*, 28 (3), 14–23.

Nicol, D. J. and Macfarlane-Dick, D. (2006) Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education,* 31 (2), 199–218.

Nuthall, G. (2007) *The hidden lives of learners.* Wellington: Nzcer Press.

Osterlind, S. J., (1998) *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance and Other Formats.* Netherlands: Springer

Pellegrino, J. W., Chudowsky, N., and Glaser, R. (eds.) (2001) *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.

Pellegrino, J. (2014) A Learning Sciences Perspective on the Design and Use of Assessment in Education. In R. Sawyer (ed.), *The Cambridge Handbook of the Learning Sciences*. Cambridge: Cambridge University Press.

Pellegrino, J.W., DiBello, L. V. and Goldman, S. R. (2016) A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist*, 51 (1), 59–81.

Popham, W. J. (2003) *Test better, teach better: The instructional role of assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Popham, W. J. (2008) Formative assessment: Seven stepping stones to success. *Principal Leadership*, (September), 16–20.

Popham, W. J. (2009) Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48 (1), 4–11.

Roediger, H. L. and Butler, A.C. (2011) The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15 (1), 20–27.

Sadler, D. R. (1998) Formative Assessment: revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5 (1), 77–84.

Shepard, L. A. (2007) The Role of Assessment in a Learning Culture. *Educational Researcher*, 29 (7), 4–14.

Shepard, L. A. (2008) A brief history of accountability testing, 1965–2007. In K. E. Ryan , L. A. Shepherd (Eds.) *The future of test-based educational accountability*. 25-46.

Shepard, L. A. (2009) Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice*, 28 (3), 32–37.

Shepard, L. A., Penuel, W. R. and Davidson, K. L. (2017) Design principles for new systems of assessment. *Phi Delta Kappan*, 98 (6), 47–52.

Shepard, L. A., Penuel, W. R. and Pellegrino, J. W. (2018a) Classroom assessment principles to support learning and avoid the harms of testing, *Education measurement: Issues and Practice*, 37 (1), 52–57.

Shepard, L. A., Penuel, W. R. and Pellegrino, J. W. (2018b) Using learning and motivation theories to coherently link formative assessment, grading practices and large-scale assessment. *Educational Measurement: Issues and Practice*, 37 (1), 21–34.

Skinner, B. F. (1985) Cognitive science and behaviourism. *British Journal of Psychology*, 76 (3), 291–301.

Smeyers, P. (2016) Neurophilia: Guiding Educational Research and the Educational Field? *Journal of Philosophy of Education*, 50 (1), 62–75.

Soderstrom, N. C. and Bjork, R. A. (2015) Learning Versus Performance: An Integrative Review. *Perspectives on Psychological Science*, 10 (2), 176–199.

Stewart, W. (2012) Think you've implemented assessment for learning? [Online]. *Times Educational* Supplement, 13 July. Available from https://www.tes.com/news/think-youve-implemented-assessment-learning [Accessed 29 August 2018]

Stobart, G. (2008) *Testing times: the uses and abuses of assessment.* Oxon: Routledge.

Stobart, G. and Eggen, T. (2012) High-stakes testing - value, fairness and consequences. *Assessment in Education: Principles, Policy and Practice*, 19 (1), 1–6.

Stobart, G. (2014) *The expert learner: challenging the myth of ability*. Berkshire: Open University Press.

Swaffield, S. (2009) The misrepresentation of Assessment for Learning – and the woeful waste of a wonderful opportunity. [Online]. In: *AAIA National Conference,* Bournemouth, 16-18 September 2009. Available from: https://www.aaia.org.uk/content/uploads/2010/07/The-Misrepresentation-of-Assessment-for-Learning.pdf [Accessed 29 August 2018]

Swaffield, S. (2011) Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy and Practice*, 18 (4), 433–449.

Torrance, H. (2007) Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy and Practice*, 14 (3), 281–294.

Wiliam, D. (2006) Assessment for learning: why, what and how. [Online]. In: *Cambridge Assessment Network Conference*, Cambridge*,* September 15 2006. Available from: https://www.dylanwiliam.org/Dylan_Wiliams_website/Papers.html [Accessed 29 August 2018]

Wiliam, D. (2011a) *Embedded formative assessment.* Bloomington: Solution Tree Press.

Wiliam, D. (2011b) What is assessment for learning? *Studies in Educational Evaluation*, 37 (1), 3–14.

Wiliam, D. [DylanWilliam] (2013, 23 October) Example of really big mistake: calling formative assessment formative assessment rather than something like "responsive teaching" [Tweet]. Retrieved from https://twitter.com/dylanwiliam/status/393045049337847808?lang=en

Wiliam, D. (2014) Formative assessment and contingency in the regulation of learning processes. [Online]. In: *Toward a Theory of Classroom Assessment as the Regulation of Learning*, Philadelphia, April 2014. Available from: https://www.google.com/search?client=safari&rls=en&q=Formative+assessment+and+conti

ngency+in+the+regulation+of+learning+processes.&ie=UTF-8&oe=UTF-8 [Accessed 29 August 2018]

Wiliam, D. (2016) *Leadership for teacher learning*. West Palm Beach, FL: Learning Science International.

Wiliam, D. (2017) Learning and assessment: a long and winding road? *Assessment in Education: Principles, Policy and Practice*, 24 (3), 309–316.

Wiliam, D. (2018) How Can Assessment Support Learning? A Response to Wilson and Shepard, Penuel, and Pellegrino. *Educational Measurement: Issues and Practice*, 37 (1), 42-44.

Wiliam, D. and Thompson, M. (2007) Integrating assessment with instruction: what will it take to make it work? In: C. A. Dwyer (ed.), *The future of assessment: shaping teaching and learning* (53-82). Mahwah, NJ: Lawrence Erlbaum Associates.

Willingham, D.T. (2008) When and How Neuroscience Applies to Education. *Phi Delta Kappan*, 89 (6), 421–423.

Willingham, D. T. (2009) Three problems in the marriage of neuroscience and education. *Cortex*, 45 (4), 544–545.

Wilson, M. (2018) Making measurement important for education: The crucial role of classroom assessment. *Educational Measurement: Issues and practice*, 37 (1), 1–37.

Wylie, E. C. and Lyon, C. J. (2015) The fidelity of formative assessment implementation: issues of breadth and quality. *Assessment in Education: Principles, Policy and Practice*, 22 (1), 140–160.

Zimmerman, B. J. (2000) Attaining self-regulation: A social cognitive perspective. In M. Boekaerts and P. R. Pintrich (eds.). *Handbook of self-regulation.* Diego, CA: Academic Press, 13-39.

Zimmerman, B. J. (2002) Becoming a Self-Regulated Learner: An Overview. *Theory into practice*, 41 (2), 64–70.

Zimmerman, B. J. and Moylan, A. R. (2009) Self-regulation: Where metacognition and motivation intersect. In Hacker, D. J., Dunlosky, J. and Graesser, A. C., (Eds.) Handbook of metacognition. New York: Routledge

Appendix A – Letter of permission to Headteacher

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

[redacted]

[redacted]

Dear [redacted],

I am writing to enquire about conducting research in school this academic year. As you know, I am studying for the Master's in Learning and Teaching at Oxford University, supervised by [redacted]. In my final research project 'A cognitive approach to assessment for learning; evaluating the impact of changing the nature and timing of formative assessment in a secondary school science faculty.' I will attempt to evaluate the impact on pupil attainment and on the learning process.

The changes that we have implemented in faculty and impact of these will be assessed using a mixed-methods approach. Attainment data will be collected at the three assessment points during the school year and these, along with teacher and pupil interviews and questionnaires will be used to evaluate the impact on pupil attainment and the impact it has made on the learning process. The project will also compare the attainment data to that of the previous academic year and teachers and pupils will be asked as part of their interviews/questionnaires to compare the new practices to the previous ones.

By participating in the research, the school would be contributing to a project that will consider the impact of assessment on the learning of our students and I hope this will bring benefit not just to the science faculty but also to other areas of the school if suitable. I hope to conduct this research between November and July. I would give questionnaires to students and interview the science teachers who have been involved in the implementation of the new assessment systems.

Oxford University has strict ethical procedures on conducting ethical research, consistent with current British Educational Research Association guidelines. The University also recognises, however, that my study is a piece of practitioner research, and that schools already operate with the highest ethical standards. Therefore, only your formal consent as head teacher is necessary, and not that of individual parents or staff. However, throughout the research, students and other teachers will be able to refuse to participate in any research activities at any time.

All participants, including students, teacher and the school, would be made anonymous in all research reports. The data collected would be kept strictly confidential, available only to my supervisor ([redacted]) and me, and only used for academic purposes. It will be kept for as long as it has academic value.

If you are happy for me to proceed with this study, please confirm that using the attached reply form. If you have any concerns or need more information about what is involved, please contact me or my

supervisor. Further, if you have any questions about this ethics process at any time, please contact the chair of the department's research ethics committee, though: research.office@education.ox.ac.uk

I look forward to hearing from you.

Yours sincerely,
[redacted]

A cognitive approach to assessment for learning; evaluating the impact of changing the nature and timing of formative assessment in a secondary school science faculty

[redacted]

☐ We do not wish to participate in this project.

☐ We would like to find out more about this project.

☐ We would like to take part in this project.

_____

Head teacher's signature

Please return this form to me.

Appendix B – Headteacher's signed consent

A cognitive approach to assessment for learning; evaluating the impact of changing the nature and timing of formative assessment in a secondary school science faculty

[redacted]

☐ We do not wish to participate in this project.

☐ We would like to find out more about this project.

☒ We would like to take part in this project.

[redacted]

Head teacher's signature

Please return this form to me.

Thank you for your help

Appendix C – Example of student survey with coding

(4)

Science Student Survey

Thank you for taking part in this survey. Please answer questions honestly and take it seriously. If you are not sure what a question is asking, then please ask your teacher.

Please be aware that some of your answers may be used anonymously in a research project that [redacted] is currently completing. If you would like your answers to be withdrawn from this project, please let [redacted] know.

Year and Class: [redacted]

Please circle a number for the following statements. 1 = strongly disagree, 2 = disagree, 3 = agree and 4 = strongly agree.

	Strongly disagree	Disagree	Agree	Strongly agree
Doing progress checks has helped my learning	1	2	(3)	4
I know how to improve after doing a progress check	1	2	3	(4)
I understand why we do the progress checks in science	1	2	3	(4)
I think progress checks are useful	1	2	(3)	4
I use progress checks to help me revise	1	(2)	3	4
Progress checks help me do better in exams	1	2	(3)	4
I would rather do progress checks than end of topic tests	1	2	3	(4)

Please explain your answer to the **last** question above.

I would rather do progress checks because we get instant feedback back whereas end of topic test teachers have to collect them then mark all of them.

(H)

PLEASE TURN OVER

What do you **most like** about progress checks? Think about your learning.

I like the fact that sometimes we get instant feedback which in my opinion helps me see where I went right & wrong.

What do you **not like** about progress checks? Think about your learning.

Not being told when we will be doing one.

How would you like to see progress checks changed in the future?

Progress checks have replaced end of topic tests in science. What is the difference between these? (Please only answer if you can, if you are not sure leave this blank).

Thank you for completing this survey

Appendix D – Example section of analysis from student surveys

										Explain your answer to Q7			What do you most like about progress checks? Think about your learning.			What do you not like about progress checks? Think about your learning.		
Year	Class	Survey Number	Q1	Q2	Q3	Q4	Q5	Q6	Q7	1	2	3	1	2	3	1	2	3
[redacted]	[redacted]	1	2	3	3	1	1	2	2	X			A			G	F	
		2	2	3	4	3	2	2	2				C			X		
		3	3	3	4	4	3	3	3	I	J		A					
		4	3	3	4	4	2	3	4	I	J		A	B				
		5	2	3	2	2	2	2	3	A			A			I	E	
		6	3	3	3	3	2	3	3	I	D	K	A			F	H	
		7	3	3	4	3	3	3	3	J			A					
		8	3	3	4	4	4	3	3	I	J		B					
		9	3	4	4	3	2	4	3	I			A					
		10	3	3	4	3	3	3	4	L			D	B				
		11	3	3	3	3	2	3	3	J								
		12	3	2	3	3	2	3	4	D								
		13	3	3	3	3	2	3					E			H		
		14	3	2	4	3	4	3	3	D	I		B	F		E		
		15	3	3	3	3	3	3	4				G					
		16	3	4	3	3	2	3	4				A					
		17	3	3	3	3	2	3	3	J								
		18	3	4	3	4	4	4	4	C	I	D	A	B		X		
		19	3	3	3	3	2	2	3				H			I		
		20	3	4	4	3	3	4	4				A	F		I		
		21	3	4	3	4	3	3	3				A	B		E		

Appendix E – Student and teacher interview questions

Student interview questions

1. How do progress checks affect your learning (if at all) before you do them?
2. How do progress checks affect your learning (if at all) as you do them?
3. How do progress checks affect your learning (if at all) after you do them?
4. How do teachers use the progress checks?
5. How do progress checks compare to the end of topic tests that we used to do?
6. What is the purpose of the review lessons that follow the progress checks?
7. What are the advantages of progress checks?
8. How could progress checks be changed or improved further?

Teacher interview questions

1. What, in your experience, is the purpose of the progress checks?
2. What impact has the intervention made to your pupils' learning? How do you know this?
3. Has there been an impact on pupil achievement?
4. What impact has the intervention made to your teaching?
5. What are the consequences of using the progress checks as opposed to the previous end of topic tests?
6. If timing has not already come up, what are your views on the timing changes?
7. Has the intervention had a positive impact?
8. How do you know this?
9. Are their negative impacts of the intervention?
10. What are these?
11. How would you like to see the intervention developed in the future?

Appendix F – Example sections of interview transcripts with analysis

Student Interview

Interviewer: The first three things I want to ask you about are how the progress checks affects your learning, if so how, before you do them, as you do it and then after you do the test. Shall we start with as you do it?	
Student: Yeah, I think, as I'm doing it I kind of notice the gaps that I've, in my learning that I've um.. as I'm doing it I kind of recognize what kind of questions I'm not really familiar with and then I can apply that into maybe, like, revision further... I know I need to work on that. Or when the teacher marks it after as well I see, like, the targets and where I need to develop my answers.	*Identifying strengths and weakness* *Use for revision* *Helpful for improvement*
Interviewer: So, as you're doing it you see for yourself gaps. Do you know where you've got things right and wrong?	
Student: Yeah, it's like, I can automatically tell reading the question.	*Questions themselves as diagnostic*
Interviewer: Just reading the question you know whether you'll get it right?	
Student: Yeah	
Interviewer: So... let's go to afterwards, how is your learning benefitted or not benefitted by having done the progress check?	
Student: Well, as I said before I can see where I went wrong and where I need to develop and that kind of, I kind of learn from my mistakes and its kind of, I feel more comfortable doing a progress check than an end-of topic test... I don't know if there's less pressure or something.	*Developmental value/diagnostic* *More comfort/less stress*

Teacher Interview

Interviewer: So, first question is, what in your experience is the purpose of the progress check?	
Teacher: The progress check tells me how much the children have learnt in that two week period that I've been teaching them. Um... not only does it tell me what they've learnt but I kind of use it as, it tells me how good my teaching was. So if I thought, right, great, I've taught this bit perfectly, you guys should all get it right... and I look at the progress checks and no-one's got it right I think, OK, I've done that bit rubbish. So, it helps me to see where they're at, and how I'm doing with teaching the content.	*Use for teacher – what students have learnt* *Use for teacher – how well they have taught*
Interviewer: If we think about the pupils now, I want you to think and reflect on what impact doing the progress checks makes to your pupils learning.	
Teacher: Um... the kids tend to take the lessons very seriously with the idea in mind that they want to do well in their progress check. So instead of them waiting until the end of the cycle or the end of... until the end of cycle tests... they actually, I hear them thinking 'oh, miss, is this going to be in the progress check?' or they put their hand up and ask 'can I take this home to revise for the progress check?' So I think they've always got the progress check in their mind because they want to be able to show what they've learnt, and then when they've done well on the progress check they are really proud of themselves. 'Cause the first thing, when I give it back to them, they turn straight to the box and see 'what's my target?' or 'how many things do I need to work on?' 'cause I make a big deal about that box.	*Improvement in motivation?* *The assessment leading the curriculum* *Pride in their work – motivation* *Identifying targets*
Interviewer: Yep	
Teacher: So, yeah, they kind of learn with it at the forefront of their mind and they look forward to getting it back.	

Appendix G – Years 7–10 Target data

	Year 7	Year 8	Year 9	Year 10	Year 11
% Target 1+	100	100	100	100	100
% Target 2+	99	99	100	100	100
% Target 3+	96	92	94	97	92
% Target 4+	75	81	79	90	73
% Target 5+	49	55	46	67	35
% Target 6+	25	25	21	35	18
% Target 7+	6	8	4	17	4
% Target 8+	1	1	1	6	0