Formative use of summative assessment in Year 7 Science.

Xarius Austin

A Research & Development Project

Submitted for the MSc Learning & Teaching 2021

DEPOSIT AND CONSULTATION OF THESIS

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

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

Surname	Austin
First Name	Xarius
Faculty Board	Education
Title of Dissertation	Formative use of summative assessment in Year 7 Science.

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 Oxford Research Archives 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.

Formative Use of Summative Assessment in Year 7 Science

Table of Contents

Introduction

The national curriculum in England is currently based around an extensive programme of national assessment, dating back to the start of the 1990s (Green & Oates, 2009). In its current form, students in England are subjected to summative, external assessments at the end of Key Stage 2 and 4. At the end of Key Stage 4, most students take a multitude of GCSEs in English, maths, science and a range of other compulsory and elective subjects. The grades they achieve in these exams are pivotal to each students' career, whether they go on to further study or into employment. Since the removal of science as an externally assessed subject at the end of KS2 as a part of the SATS exams in 2009 (Curtis, 2009), GCSE science exams will be the first external summative examination that students will have sat in science in their school career. As a secondary school science teacher, I wanted to explore how we prepared students for those examinations throughout their time at secondary school. The inspiration for this project came during a science faculty meeting when we were discussing our Year 7 scheme of work, and whilst talking about it to an experienced science teacher, she raised the valid observation that we test Year 7s widely but never consistently teach them how to tackle the exam questions in the first place.

School Context

My school is a mixed comprehensive academy, based in Oxfordshire. The school is unusually large, with 2200 students on roll, and approximately 320 being entered for GCSE exams each year. 6.1% of children are eligible for free school meals. All students sit 'Combined Science' (formerly known as the 'Double Award'), gaining two GCSEs in science at the age of 16. This is different to the majority of schools, who usually offer separate sciences (formerly known as 'Triple Science') in addition to 'Combined Science'. With between 320-350 students in each year, our science department is considerably large, with nearly 30 teachers, most of whom are part-time. In Year 7, students have 2 hours and 10 minutes of science each week, usually taught by one teacher. In years 8 and 9, this increases to 3 hours and 15 mins each week, again usually taught by one teacher. In years 10 and 11, this again increases to 4 hours and 20 mins per week, taught by two teachers cooperatively. The teachers usually have different science specialisms.

Primary School Experiences

Students enter Year 7 from up to 28 different primary schools, with 90% of students coming from 10 core primary schools. Students join the school with a huge range of prior science knowledge, because different primary schools place differing levels of importance on science given that the subject does not feature in Key Stage 2 external assessments (Department for Education, 2014), and by extension, does not contribute to the scoring of their primary school in league tables. A report commissioned by the Wellcome Trust found that over half of primary school classes do not receive two hours of weekly science, with 12% of primary schools not delivering any weekly science lessons to any year groups (CFE Research, 2017). Students also have varying experiences with assessment, both formative and summative at primary, as this is largely determined by each individual school, however all students will have taken their SATs exams in English and mathematics. This disparity in prior learning and assessment means that our Year 7 curriculum has been designed to make sure all students are not disadvantaged depending on the primary school they come from, by teaching students in mixed ability tutor groups, and assuming no prior learning, as well as starting all students in Year 7 on the same topic.

Current Assessment for Learning Practice

Currently, during Key Stage 3, all students are taught six science modules (2 of each specialism; Biology, Chemistry, and Physics) in mixed-ability tutor groups. They sit an internally assessed summative topic test at the end of each module, and an internally assessed summative end-of-year test. Students are currently given minimal and inconsistent support when it comes to teaching them how to understand the language of exam questions. It is not written into the scheme of work, therefore the training students are given with regards to exam practice is dependent on the teacher, with most admitting that they do not focus on this at Key Stage 3. The end-of-year test in Year 9 is internally moderated as it is used to determine places in our two top sets at GCSE level. This replaced the Year 9 SATs exams that were abolished in 2008 (Curtis, 2008). All other students are placed into mixed ability groups at Key Stage 4. The results of these tests are used to inform three student reviews that occur each year. During these reviews, parents are told whether their child is performing ‘Above Target’, ‘On Target’ or ‘Below Target’. These targets are

determined by student's performance in Key Stage 2 SATs, combined with the results of Cognitive Abilities Tests (CATs) sat within the first month of Year 7. These results are then interpreted by FFT, a not-for-profit part of the Fischer Family Trust, who take this data and match it to students from schools across the country to predict the most likely grade they will achieve at GCSE level (FFT, n.d.).

In years 10 and 11, students have 4 hours 20 minutes of science each week, taught by two teachers, usually with different specialisms. Throughout Key Stage 4, students are assessed in a similar way to Key Stage 3. They learn six modules per year, sitting an internally assessed summative topic test, as well as one end-of-year internally assessed summative test in Year 10, and one internally assessed summative mock exam in Year 11. These tests are created from past paper questions. At Key Stage 4, we focus more about the language of exam questions in these topic tests, and throughout topics as practice exam questions have been written into the scheme of work in most lessons. All students are given a card which lists the different exam question 'command words' as set out by the exam board in every test, however the use of them by students in a test situation still seems fairly limited. The results of these tests are also used to inform three review periods. In Year 11, students also sit two science 'Walking Talking Mock' style exams. The concept was initially made popular in England by The PiXL Club – Partners in Excellence, who are an apolitical network of over three thousand schools working together to share best educational practice and raise standards (PiXL, n.d.). A 'Walking Talking Mock' consists of students sitting a mock exam under exam conditions, however a teacher leads the pupils through the exam, highlighting what to look for to decipher the language of the exam questions, modelling what a student should do in the real exam, giving exam technique advice whilst students sit the exams (Wilson, 2019).

Project Aims

We focus a lot of our lesson time in years 10 and 11 on teaching 'exam technique', which are the skills to decipher the language of exam questions to figure out what an exam question is looking for. I feel that as students have been tested so vigorously in Year 7, 8 and 9, a lot have embedded bad exam habits which are difficult to break in Year 10, so I wanted to explore whether instilling good 'exam technique' skills from the start of Year 7 made a

difference to their overall performance in their GCSEs. As this project only lasted one year, I felt the best way to investigate this was by teaching a sample of classes these skills, using summative assessment in a formative way when they would usually sit their first couple of topic tests, and comparing their performance in the subsequent summative topic tests and end-of-year test to classes that have not been taught these skills. To teach these skills, I wanted to explore if the 'Walking Talking Mock', as students sit two of them in Year 11, was the right vehicle for teaching those exam skills, and whether focusing on understanding the language of exam questions improved attainment. My trial aimed to explore whether the academic data showed an increase in attainment when using summative assessment formatively compared to the control groups. I also felt interviewing experienced teachers taking part in the trial was necessary to gain feedback on whether they felt that further implementation of this would increase future attainment. Finally, I wanted to gauge student feedback about the trial. If these showed positive results, we would aim to implement this into our scheme of work for all Year 7 classes in the following academic year, even possibly extending to Year 8 and Year 9 depending on the results.

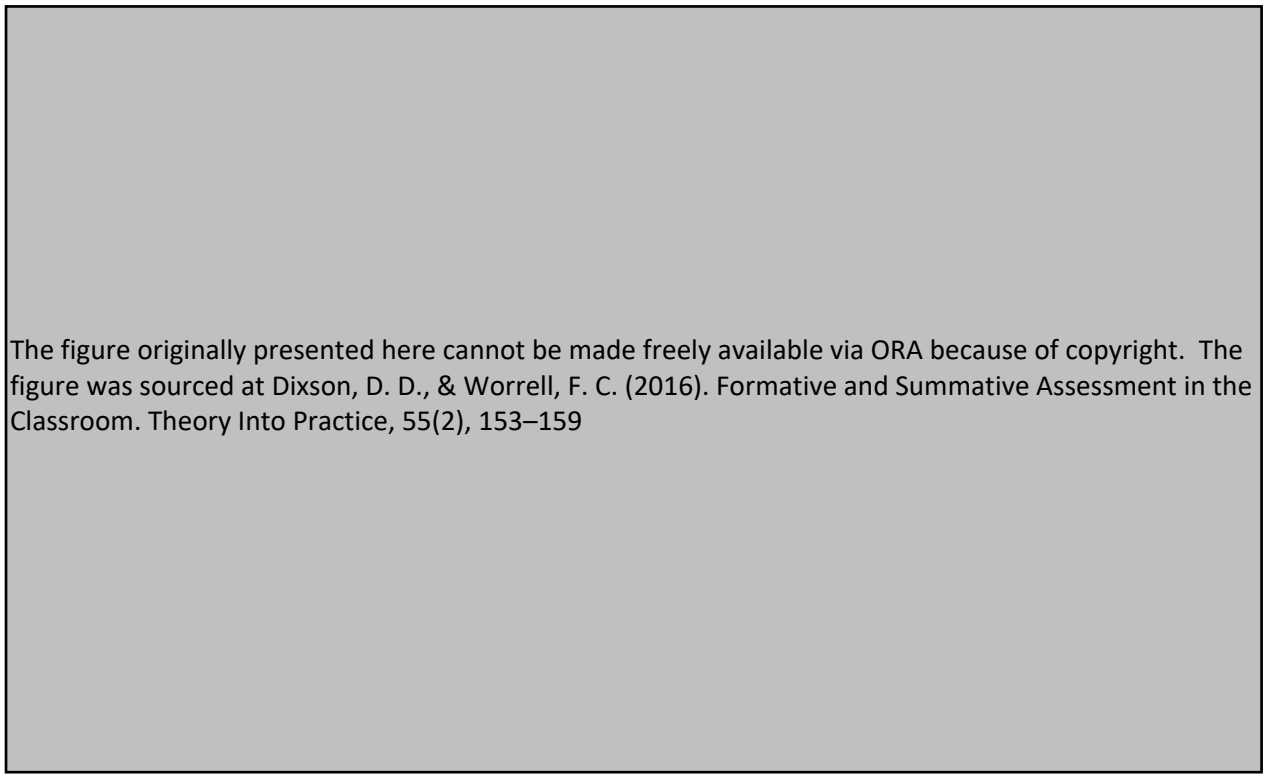
I have chosen to adopt a narrative style approach to explain my research. Adoption of a narrative research style is progressively becoming more widespread when telling stories of educational practice and experience (Moen, 2006). As this research is telling a story of my approach to this research, and the qualitative experiences of the students and teachers taking part, I aim to follow the guidelines of narrative inquiry set out by Connelly & Clandinin (1990) in order to enable the reader to understand and relate to my thought process and decision-making (Connelly & Clandinin, 1990).

To conduct this inquiry, in my literature review I aim to look into the background of the assessment-based model of the English education system to explore whether our reliance on examination results means teaching 'exam technique' is a time-consuming problem faced solely by English teachers, or more widespread. I want to investigate the impact such widespread testing has on our students, in order to gain an understanding of the benefits and consequences of testing to the learner. I then aim to define the benefits and challenges of both formative and summative assessment. Finally, exploring how the teaching of 'exam technique' and understanding the language of exams is essential to understanding the background to this problem.

Critical Review of the Literature

Formative and Summative Assessment

It is necessary here to clarify exactly what is meant by the terms 'formative', 'summative' and 'assessment'. Within the field of education, 'assessments' are also commonly referred to as 'tests' or even as 'evaluations' (most common in US research). Taras defines assessment as a "judgement which can be justified according to specific weighted set goals, yielding either comparative or numerical ratings." (Taras, 2005, p.467). This definition is clear, but has a very narrow scope. Black, however, defines assessment as "the process by which teachers use learners' responses to specially created or naturally occurring stimuli to draw inferences about the learners' knowledge and skills" (Black et al., 2004, p.2). Black's definition is useful because it takes into account both the teacher and learner's role in assessment, as well as identifying that inferences are made from those tests about knowledge and skill. The origin of the terms 'formative' and 'summative' assessment can be traced back to Scriven in the late 1960s. In his book, he describes formative assessment as "typically conducted during the development or improvement of a program… and it is conducted, often more than once, for the in-house staff of the program with the intent to improve." (Scriven, 1981, p.168-169). He then goes on to define summative assessment as "conducted after completion of the program… and for the benefit of some external audience or decision-maker…" (Scriven, 1981, p.340). Dixson & Worrel neatly summarise the similarities and differences of formative and summative assessments, shown in ***Table 1***. As formative assessment is usually used within schools to guide pupil improvement, it is considered low-stakes. Different examples of formative assessment can be seen throughout the teaching of a module in order to gauge student understanding. Summative assessments are usually more formal exams sat at the end of a module to evaluate learning and are considered high-stakes as the results of the assessments have implications not only for the learner, but the teacher, faculty, and school (Dixson & Worrell, 2016).



Assessment in the English Education System

Assessment in the English education system has always been multi-faceted in its functions. Most obviously, assessment serves as a way of establishing competency and knowledge of those being assessed. It serves as a benchmark for further education and higher education institutions to allocate places to prospective students. It also serves to hold teachers and schools to account for their quality of teaching, shown by the reliance on assessment data in school league tables, as well as when performance-managing teachers, and its use during Ofsted inspections (Wall, 2000).

Our current reliance on assessment data can be traced back to the late 1980s when Thatcher's Conservative government overhauled the much-criticised education system and introduced the precursors to our current education system within the Education Reform Act. This act, given royal assent in 1988, outlined plans for a National Curriculum and National Curriculum Assessment framework, as well as introducing the idea of key stages 1-4 and years 1-11 (*Education Reform Act 1988*). In his article, Whetton (2009), outlines the journey that assessment has taken from the Education Reform Act up to the 2008 unexpected removal of key stage 3 assessment. As someone who has worked at the

National Foundation for educational research, Whetton comprehensively summarised the key changes that led from the initial complicated and confusing teacher-assessment based model of assessment proposed by the Task Group for Assessment and Testing to the externally moderated examination model of assessment we use today. Whetton attributes this shift in the most part to the increasing political need to hold schools accountable for the quality of their teaching. It is important to note that Whetton himself acknowledges that given his longevity in the field, there may have been a degree of bias within his account of events. The Task Group for Assessment and Testing was predominantly made up of prominent educationalists of the time. At the helm was Paul Black, one of the most prolific researchers into formative assessment in the classroom globally.

In order to re-design the National Curriculum Assessment model, they recommended that their new assessment model should meet 5 key criteria:

- *formative – providing information on where a pupil is, enabling teachers to plan the next stages;*
- *summative – providing overall information on the achievement of pupils;*
- *evaluative – providing aggregated information on classes and schools to assess curriculum issues, as well as the functioning of teachers and schools;*
- *informative – providing information to parents about their own children and general information about the school;*
- *for professional development – giving teachers greater sophistication in assessment, recording and monitoring so that they can evaluate their own work.*
  (Whetton, 2009, p.141)

Their proposed model, based on a 10-level structure of teacher assessment was not well received politically, with Thatcher herself calling it a “a weighty, jargon filled document” (Thatcher, 1993, p. 594) in her memoirs.

What followed were years of confusion and unrest by teachers, escalating to key stage 3 English teachers and unions threatening to boycott the 1993 examination series, leading to Ron Dearing, a senior civil-servant who was formerly Chief Executive of the Post Office, and at the time, Chancellor of the University of Nottingham, being called upon to reform the assessment model in 1994. The Dearing review highlighted that it needed to focus on:

> *i reducing the volume of material required by law to be taught;*
>
> *ii simplifying and clarifying the programmes of study;*
>
> *iii reducing prescription so as to give more scope for professional judgement;*
>
> *iv ensuring that the Orders are written in a way which offers maximum support to the classroom teacher*
>
> (Dearing, 1994, p. 17)

Whetton hailed Dearing as the saviour of education, however Gillard (2018) acknowledged that not all felt the same way.

> *Myra Barrs, Director of the Centre for Language Education, argued that Dearing had left a 'trail of confusion' behind him, which was 'likely to be most marked at Key Stage 4', where he had 'casually demolished a common system of examining at 16+ which it had taken more than twenty years to establish'*
>
> (Gillard, 2018)

By 1995, summative tests in all key stages were being marked externally. With Labour's resounding victory in 1997, came more change to education. In order to measure meaningful changes to the education, the Labour government moved to make the assessment system more target driven, stating that all students should be achieving a Level 4 by the age of 11.

By 2005, teachers and parents were arguing that rigorous externally examined assessment was having a detrimental effect on the wellbeing of students. After a review, the key stage 1 assessments moved back to a teacher-assessed model, with external examinations still facing students at key stages 2, 3 and 4.

In 2008, another radical move was made by secretary of state, Ed Balls. He put an end to the external assessments at key stage 3 without warning, and the following year, removed science from the key stage 2 examination requirement (Whetton, 2009).

With Labour being defeated by a Conservative/Liberal Democrat coalition in 2010, Michael Gove was made Secretary of State for education. He ordered a review of the National Curriculum and its assessment, opting for a more fact-based approach to learning, focusing overwhelmingly on content, rote learning and external examinations, removing a number of vocational and coursework-based courses (Gillard, 2018). This curriculum model based around high-stakes testing is being followed in England at the time of writing.

When considering why this intervention is necessary, it is important to consider this history behind the assessments that students face in the English education system. George Madaus, a Boisi Professor Emeritus of Education and Public Policy at Boston College, has been a vocal critic of high-stakes testing. He believed that high-stakes testing corrupted the very social processes that they were due to monitor. He explained that because the results of the standardised testing have become so critical for both students and teachers, that teaching has been extraordinarily distorted towards test preparation. Indeed, he is critical of the very issue this research project aims to explore. However, Madaus notes that in reality, if the test is crucial for not only a teacher's progressions, but to student progression and career goals, undoubtedly teachers will teach to it to satisfy the superficially observable demands of teaching (Madaus, 1998).

Impact of assessment on students

To understand the impact of this intervention on students, it is important to consider the research on the impact that assessment has on students generally. As the planned intervention aims to prepare students better for the assessments they face in year 7, students will be asked for feedback as to their views on the effectiveness of the intervention in the broadest senses.

High-stakes testing refers to assessment where the results affect multiple stakeholders. In this case, student performance not only impacts on the student, but their teachers, the school, the local authority and the national picture of education (Zhao et al., 2016). Not only can the GCSE exams that students face in year 11 be considered a high-stakes test, but the

end of module tests we carry out can also be considered high-stakes tests. This is because the results of the end of module tests are used to monitor student progress throughout science within each school year, as well as within teacher performance management conversations. During the coronavirus pandemic, the results of these tests were also used to determine Centre Assessed Grades when external examinations were cancelled in 2020 and then again in 2021 to determine the Teacher Assessed Grades.

Dr Seuss famously highlighted the issue of high-stakes testing in his 1998 book 'Hooray for Diffendoofer Day', when the headteacher in the story says:

All schools for miles and miles around
Must take a special test.
To see who's learning such and such –
To see which school's the best.
If our small school does not do well,
Then it will be torn down,
And you will have to go to school
In dreary Flobbertown
(Seuss et al., 1998)

The impact of high-stakes testing has been well documented worldwide, at every level of education. In their thorough examination of the impact of high-stakes testing on teachers and students in North Carolina, Jones et al. (1999) surveyed 236 teachers. They found that 61% of teachers surveyed felt that high-stakes testing increased anxiety levels in students due to the stress of the anticipation of the test, as well as the emotional response whilst completing a test. The study would have been more useful if the authors had considered student perceptions of testing as well, however it must be acknowledged that the study focused on elementary school teachers, so insights from students in those age ranges may have proved challenging to obtain.

Madaus & Clarke (2001), prolific researchers into the impact of high-stakes testing, take this further by exploring how minority students are disproportionately affected by high-stakes testing in a number of ways. Firstly, the authors speak of the impact of high-stakes testing on motivation. Motivation has often been a common feature of arguments supporting the use of high-stakes testing dating back to the 18th Century; students are motivated by high-stakes testing to study and improve in order to do well, especially within a competitive environment. Researchers into motivation have long hypothesised that conceptual change occurs when a student's motivational beliefs are goal oriented, related to the distinction

between belief and opinion, or have perceived interest or value attached (e.g. Pintrich, 2000; Pintrich et al., 1993; Pintrich & De Groot, 1990). Yin et al. (2008) conducted a small study of 12 middle school classes to explore the impact of assessment on student motivation. Six groups had embedded formative assessment through their curriculum, and the remaining control groups were taught the curriculum as normal, which did not contain any formative assessment. They found no statistically significant increase in motivation, measured using well-established motivation questionnaires, in the experimental groups (Yin et al., 2008). Madaus & Clarke analysed 100 years of test data to conclude that the unmotivated are not motivated by high-stakes testing. Those students that regard the examinations either unobtainable, unnecessary, or feel they lack the ability to do well in them, immediately disregard them (Madaus & Clarke, 2001). This is supported by the conclusions of Minarechova (2012) who stated that one of the negative effects of high-stakes testing was that students became increasingly frustrated which led to a subsequent reduction in effort.

Minarechova (2012) also argues that high-stakes testing encourages a gender disparity in education, with teachers treating girls who underperformed in a test differently to their male classmates. Females were encouraged to re-sit and given constructive criticism in order to help them achieve their best, whereas boys were identified as unruly or not having worked hard enough. The study that Minarechova discusses, by Booher-Jennings (2008), also found that parents reacted differently when girls and boys failed tests. Girls were worried about disappointing their parents, who often encouraged to improve and do their best, whereas boys were afraid that their parents would be angry at them for failing the test. Minarechova continues to document the adverse effects of high-stakes testing on students' mental and physical health. Stress and tension in students can undermine a student's self-esteem, which can then develop into anxiety disorders, exhaustion, hypertension, gastrointestinal problems, headaches and insomnia, amongst others. This can not only lead to manifestations of erratic problematic behaviour both in school and at home, but poor concentration, nervousness and depression.

An Australian review by Polesel et al. (2012) into the impacts of high-stakes testing on school students and their families sought to examine "whether the regime of high-stakes testing throughout the school years is in the best interests of students" (Polesel et al., 2012,

p.4). The international literature review concluded that high-stakes testing negatively impacted the wellbeing of children, supporting Minarechova's summary. They documented the strain high-stakes testing put on the relationships between teachers, students and parents, often impacting a students' self-esteem and in some cases, the development of suicidal ideation related to the impact of high-stakes testing (Polesel et al., 2012). In a Canadian study, Kearns, (2011) set out to explore the perceptions of marginalised youth who failed a high-stakes test. Words used by students in the study included feeling "degraded", "stressed", "humiliated" whilst others described declining self-esteem and were considering whether they belonged in the class (Kearns, 2011).

It is clear from wider research that the impact of high-stakes testing is not only detrimental to the student's mental and sometimes physical health, but unfairly disadvantages students from minority or marginalised backgrounds.

Using Summative Assessment Formatively

There have been multiple studies exploring the link between summative assessment and student motivation, however as the concept of 'teaching to the test' is widely considered poorer pedagogy by educationalists, there is a large gap in the research into how effective different test preparation methods are. As has been already described, 'teaching to the test' is unavoidable in a system that so heavily relies on the outcomes of summative assessment, not only for student outcomes but for teacher performance management purposes. The reality of teaching in an environment that places the highest of stakes on pupil outcomes of summative tests, often with a time-pressured, content-heavy curriculum, means that teachers will undoubtedly aim to coach students to answer the types of questions found in past standardised tests (Popham, 2001).

'Teaching to the test' can take many forms, from teaching about the language of exam questions, through to the way students should approach different question types through repetitive practice by answering and marking past-paper questions. As Childs et al. (2015) explains, often the biggest barrier to student success in summative examinations is their literacy. How well a student understands the language of the exam question, and specifically the command words in the question, when reading it, often is the biggest hurdle for a student to overcome in an exam. Childs et al. advise that teachers should give out a list

of command words, similar to that found in ***Table 2*** and their explanations, and then use them throughout their teaching, reinforcing their meaning continuously (Childs et al., 2015).

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at Childs, P. E., Markic, S., & Ryan, M. C. (2015). The Role of Language in the Teaching and Learning of Chemistry. In Chemistry Education (pp. 421–446)

A newer method of preparing students for exams, that has not been the focus of any peer-reviewed research as of yet, is the idea of 'Walking Talking Mocks'. There have, however, been articles in reputable education magazines, and numerous blog posts on prolific education blogs. It seems the first time this idea surfaced in the UK was at the PiXL conference in 2012. An article in The Guardian was published about the initiative the following year (Murray, 2013). A walking talking mock aims to highlight common issues students make when answering exam questions before they answer the question in a mock-exam style setting. Students then self-mark their answers in order to understand their mistakes immediately and improve as they work through the paper (Class Teaching, 2014). A number of complementary blog posts from a range of subjects have been published subsequently (eg. English, 2018; Still, 2015). An article in the Times Education Supplement also praised the Walking Talking Mock for, anecdotally, improving GCSE grades (Cargill, 2019). One critic of the Walking Talking Mock argued that it did not improve grades in subsequent real mock exams compared to previous cohorts, however the author notes that they did boost confidence and reduce exam anxiety in his students (Robbins, 2019).

As there is a clear gap in the research into student exam preparation, this study aims to provoke a debate as to whether research like this is valid considering the consensus that 'teaching to the test' is less valued in pedagogical circles.

Methodology

Experimental Design Overview and Research Questions

The inspiration for this project came from a biology teacher in a science department meeting. Whilst discussing re-writes to our outdated Year 7 scheme of work, we were tasked with discussing how we could update the structure of each Year 7 module. I was paired with a very experienced science, biology, and child development teacher. Whilst we were talking about one of the Year 7 topic tests, she asked if I, in my revision lessons, prepared students for exam questions, in a similar way to the coaching we give to Year 10 and 11 students. I thought about it, and realised I did not… I focused on the content of the module in far greater detail and overlooked exam technique training. She told me that she, too, focused purely on the content, rather than exam technique and we started discussing the potential impact of actively teaching this at Key Stage 3, starting in Year 7. We concluded that by instilling good routines from Year 7, it may well reduce the error that arises from students not reading or interpreting exam questions properly when they reach their GCSEs and A-Level study. When choosing a project for this thesis, I remembered this conversation and I decided I would aim to explore this theory further, to see if there was any merit to the idea.

Following an initial exploration into the literature surrounding the formative use of summative assessment, I set about creating the initial research questions that would guide my experimental design. I first summarised the problem that I wanted to explore. I wanted to understand whether coaching students in exam technique skills, using summative assessment methods formatively, improved their results at GCSE. The method needed to be something we already used within our teaching repertoire so it was familiar to teachers and students would benefit from repetition, however starting it from a much earlier age than we currently start it. Due to the one-year time constraint of the project, I decided to instead focus on outcomes in Year 7, rather than outcomes at GCSE, as the latter approach would have taken five years of research and intervention. This led me to create my first research question:

1. **Does using summative assessment formatively at the start of Year 7 improve student attainment as a year group?**

I decided to use a group design method in this interventional study as it is widely regarded as the most commonly used method to evaluate the effectiveness of an intervention (Gersten et al., 2000). I achieved this by dividing the 2019 cohort of Year 7s up into two groups: the sample group, and the control group. The sample group would receive some sort of intervention, and the control group would be taught as per our current scheme of work and assessment rota. I planned to collect qualitative test data from the first four topic tests from the Year 7 cohort, as well as end of year test results. This would enable me to analyse and compare the marks of the students in the sample group with those in the control group to see if, statistically, any improvement due to the intervention had occurred.

Ideally, I wanted to select a small sample of lower, middle and higher attaining students from the sample group, control group, and previous years cohorts. I decided to include previous years cohorts in this stage to see if this year's cohort results significantly differed from previous cohorts of Year 7s. I then needed to create a way of selecting these comparable students. I would, again, use qualitative test result data. As the topic tests had not changed in the past three years, I decided to use data from 2019 (sample and control group), 2018, and 2017.

I also felt it was important to measure how the intervention specifically affected students who had differing previous attainment levels, not only in this cohort, but compared to similarly attaining students in previous cohorts as well. I felt the simplest way to judge attainment was using the data that the school based its review processes on. These include using KS2 English and Maths SATS data to predict likely GCSE outcomes, based on students with a similar educational background across the country. This analysis is carried out by a not-for-profit organisation called FFT. FFT uses a "complex formula including social and economic make-up, gender, incidence of special needs, deprivation factors, and ethnicity" (Pring & Roberts, 2015, p. 145-6) in order to create their most likely outcome at GCSE. The FFT most likely GCSE outcome is then used to create a backwards trajectory of levels

achieved in each year preceding. The school then uses their Year 7 backward trajectory level in order to base a termly review on. Students are told whether they are 'Below Target', 'On Target', or 'Above Target' in each subject they take, and this is determined by teachers, using formative and summative assessment results as a guide. Finally, Year 7 students are also given a CATS exam in their first half-term at the school. This led to me developing my second research question:

2. **Does using summative assessment formatively at the start of Year 7 improve student attainment between comparable students at KS2, CATS and FFT targets?**

Finally, as I wanted to explore the year group as a whole cohort, I needed to collaborate with other teachers in order to so. I felt it was important to take into account the views of the experienced teachers who would be teaching the sample groups. I initially settled on this research question: "Which methods did teachers find most effective to teach 'exam technique' to Year 7 students?". As described by Cobb, Confrey, DiSessa, Lehrer, & Schauble (2003), pedagogical experimental design should focus not only on crafting a method of learning, but to also explore how to, within a given educational setting, review the efficacy of the method of learning. They go on to discuss how experimental design in education needs to recognise that the aim of research is not often to test whether something empirically 'works', but to refine pedagogical methods in order to improve understanding (Cobb et al., 2003). With this in mind, I set about designing my experiment, and I decided to amend this as I did not feel exploring a range of methods to teach exam technique would have been within the scope of my research. As I wanted to mirror the methods that we already currently use in Year 10 and 11, I instead chose to focus on teacher and student perspectives of the method chosen, leading me to create my final research question:

3. **What were student and teacher perspectives about the 'Walking Talking Mock' and command word training as a way of learning/teaching exam technique?**

After designing my intervention, in order to examine my final research question "What were student and teacher perspectives about the 'Walking Talking Mock' and command word training as a way of learning/teaching exam technique?", I felt it was important to interview

the teachers taking part in the intervention, as well as quickly survey students as to their views of the intervention. Both would use qualitative data in order to gain a more subjective in-depth insight from the individuals taking part. I decided the best way to gain an in-depth insight into the views of the teachers taking part in the intervention was via an interview. I had planned this to be an audio-recorded interview in one lunchtime or after school at the teacher's convenience. In order to get an idea of student perceptions of the interview, I decided the simplest way was to get students to answer a small number of additional questions as part of their normal end of topic feedback sheet, which is at the back of every topic test.

Selecting the Sample Groups

In order to choose which Year 7 classes and teachers I would use for the sample group, I collaborated with the science department Key Stage 3 team leader after he had designed the module rota for all Year 7 classes. We looked at the module rotation shown in ***Table 3***. Ideally, we wanted to look for classes sitting a similar selection of modules for the sample group, in order to reduce the number of new resources made. We felt having a range of teachers taking part in the sample group would help reduce individual teacher bias in the sample, but we felt it might also be interesting to explore whether two classes taught by the same teacher had a similar experience.

We ended up selecting classes 1, 4, 8, 9 and 10. Initially, class 10 was chosen because they were the only Year 7 class taught by myself, and I wanted to teach one of the sample groups. All Year 7 classes start on the Particles module, however as the second module they would complete was Reactive Science, we felt including the other two classes (8 and 9) also studying Reactive Science as their second module was necessary. This gave us the options of classes 8 and 9. As both were taught by the same teacher, we did not want to confuse the teacher (T7) by making them teach class 8 something different to class 9, therefore both groups were chosen for the sample group. For the purposes of this thesis, we shall call T7 William. All names used in the thesis have been changed to preserve anonymity for the teachers taking part. William was the science department Key Stage 3 team leader, and an experienced teacher in both Australia and England. He started teaching 13 years ago and had been at the school for 7 years prior to the start of the project.

Table 3: Year 7 Teaching Rota.

The first column represents the teacher, and the second column indicates the class they are teaching as per the timetable. The order of the six Year 7 modules across the year (in abbreviated form) is then listed. The colours denote the specialism of science.

Key: P – Particles, EE – Energy & Electricity, RS – Reactive Science, MM – Medical Mysteries, FA – Forces & Astronomy, LW – Living World. Blue – Physics, Pink – Chemistry, Green – Biology.
** denotes class selected to receive the intervention and form part of the sample group.*

		1	2	3	4	5	6
T1	Class 1*	P	EE	RS	MM	FA	LW
T2	Class 2	P	EE	RS	MM	FA	LW
T2	Class 3	P	EE	RS	MM	FA	LW
T3	Class 4*	P	EE	RS	MM	FA	LW
T4	Class 5	P	MM	EE	FA	RS	LW
T5	Class 5	P	MM	EE	FA	RS	LW
T6	Class 6	P	MM	EE	FA	RS	LW
T6	Class 7	P	MM	EE	FA	RS	LW
T7	Class 8*	P	RS	MM	EE	FA	LW
T7	Class 9*	P	RS	MM	EE	FA	LW
T8	Class 10*	P	RS	MM	EE	FA	LW

We then selected class 1 and 4 based on the experience of the teacher. Drew (T1) was the current science department Key Stage 4 team leader who had been instrumental in setting up the 'Walking Talking Mock' in Year 11, and the command word training in Year 10 and 11, therefore we felt his input and insight into the resources produced and outcome of the project was invaluable. Drew had been teaching for 15 years and had joined the school 6 years prior to the start of the project. Finally, Helen (T3) was the former science department Key Stage 3 team leader who had expressed interest in being part of the sample group during a faculty meeting where I presented my project to the science department team. She

felt that she would also like to implement the proposed intervention within her own Year 8 and 9 teaching. As an experienced teacher of 15 years, with a background of a research-based PhD in Chemistry and having joined the school 7 years prior to the start of the project, we felt that her input was also going to be crucial, therefore we selected class 4.

The benefits of using this method to select the sample groups were that we had a number of very experienced teachers taking part in the intervention, as well as almost half the year group taking part in the sample, which meant my investigation had a relatively large sample size. The drawback of this method meant that many of the control groups sat a test that was not included in the sample (the medical mysteries topic). The sample also did not take into account teachers teaching within or outside their subject specialism.

Designing the Intervention

I knew that in order to get meaningful engagement from the science department, I would have to design an intervention that seamlessly integrated with our scheme of work, was easy to implement and required a minimal increase in workload as our science department was already under a considerable amount of strain, with multiple science staff members off on long term sick leave at the same time. This, coupled with a desire to repurpose the methods used to train students in this exam technique coaching done in Year 10 and Year 11 to encourage repetition and familiarity for students by the time they reach Key Stage 4, as well as familiarity for teachers who will already recognise these teaching methods from teaching Key Stage 4 classes.

I explored several options and combinations of options for the proposed intervention and settled on three that I felt were viable. I then presented the three potential interventions to Helen, who I had asked to help design the intervention with. The first idea was to add exam style questions to every lesson in the module in order to get students used to the type of exam questions. Drew, alongside the team leaders for Biology, Chemistry and Physics, had recently done this for all the KS4 modules. They inserted exam style questions that were relevant to each lesson in our scheme of work into the lesson, in order to ensure students were exposed to and became more familiar with the language of exam questions, long before their end of module test. This satisfied the requirement to use a method that teachers were familiar with, but after discussing the options with Helen, we agreed that the

time required to do this would be too intensive in an already tumultuous year. As Key Stage 3 exam style questions were not readily available, and many of our lessons cover concepts outside the national curriculum in order to foster a broad interest and love of science, many lessons would have no easily available exam questions.

The next option I explored was giving students an extra lesson dedicated to practicing exam technique when it came to revision time. The theory behind this was that we would introduce command words and how to break down an exam question before the end of module test, and then students would have sat the test as normal. A sheet of command words and their meaning would also be created, similar to the ones we use at GCSE and A Level, to help students to break down the language of exam questions, and this would be introduced in the extra lesson and also given to them in the end of module test. When discussing options, Helen and I both liked this combination of interventions, as it allowed the data collected for the first two tests to be comparable between the sample and control groups, but we felt that the extra lesson was not something teachers currently did at Key Stage 4, therefore it was less familiar and also required us collating a number of practice questions, which were not easily available at a Key Stage 3 level. We both liked the idea of a command word sheet, explaining the meaning of command words taken from Key Stage 3 exam questions, but agreed this needed to be used in collaboration with another intervention.

My final idea was taking the model of the 'walking talking mock' currently used in Year 11 a couple of months before their external GCSE examinations and adapting it to the Year 7 end of module tests. At GCSE level, the 'walking talking mock' consisted of students, under examination conditions in an examination hall with a science teacher leading the 'walking talking mock' reading the question to the students and breaking down the language used before they then answered the question in a timed exam-style scenario. When the time for the question had lapsed, the teacher would show the students the mark scheme, and students would self-mark their work. The theory behind it is that students would get instant feedback about how they are answering exam questions, rather than getting feedback days or weeks after it had been marked and returned to them, when the exam and their answers had been mostly forgotten. Helen and I both agreed that this was the most suitable intervention as it met all of my selection criteria; it was easy to create and implement,

teachers would be familiar with the concept having all taken part in them during Key Stage 4 teaching, and it integrated seamlessly with the current scheme of work. We did, however, understand that the structure and focus of the 'walking talking mock' had to be changed in order to suit a Year 7 class who were unfamiliar with science examinations. Firstly, the 'walking talking mock' would have to be done in the classroom but could still be done under exam conditions with students spaced out around the room (as is normal during an end of topic test). In addition to this, the focus on command words had to be more explicit because this was a completely unfamiliar concept to the Year 7 students. In order to scaffold this, we felt that creating and using the command word sheets during the 'walking talking mock' would achieve this. Finally, we did not want to focus on the self-marking of the test, as William had made it clear that we still needed to collect teacher-assessed data for the Year 7 reviews. We felt it would be sufficient to explain the command words and language of the question, and then let students answer the question in timed conditions, without getting the instant feedback that occurred at Key Stage 4. Instead, the teacher would mark the test and feedback to the students in their preferred way. I then spoke to William and Drew who both agreed that this seemed like the best approach to take.

Resource Design

The resources I needed to design for the intervention were:

1. Command word guidance sheet
2. 'Walking talking mock' for the Particles topic test
3. 'Walking talking mock' for the Reactive Science topic test

4. ‘Walking talking mock’ for the Energy and Electricity topic test

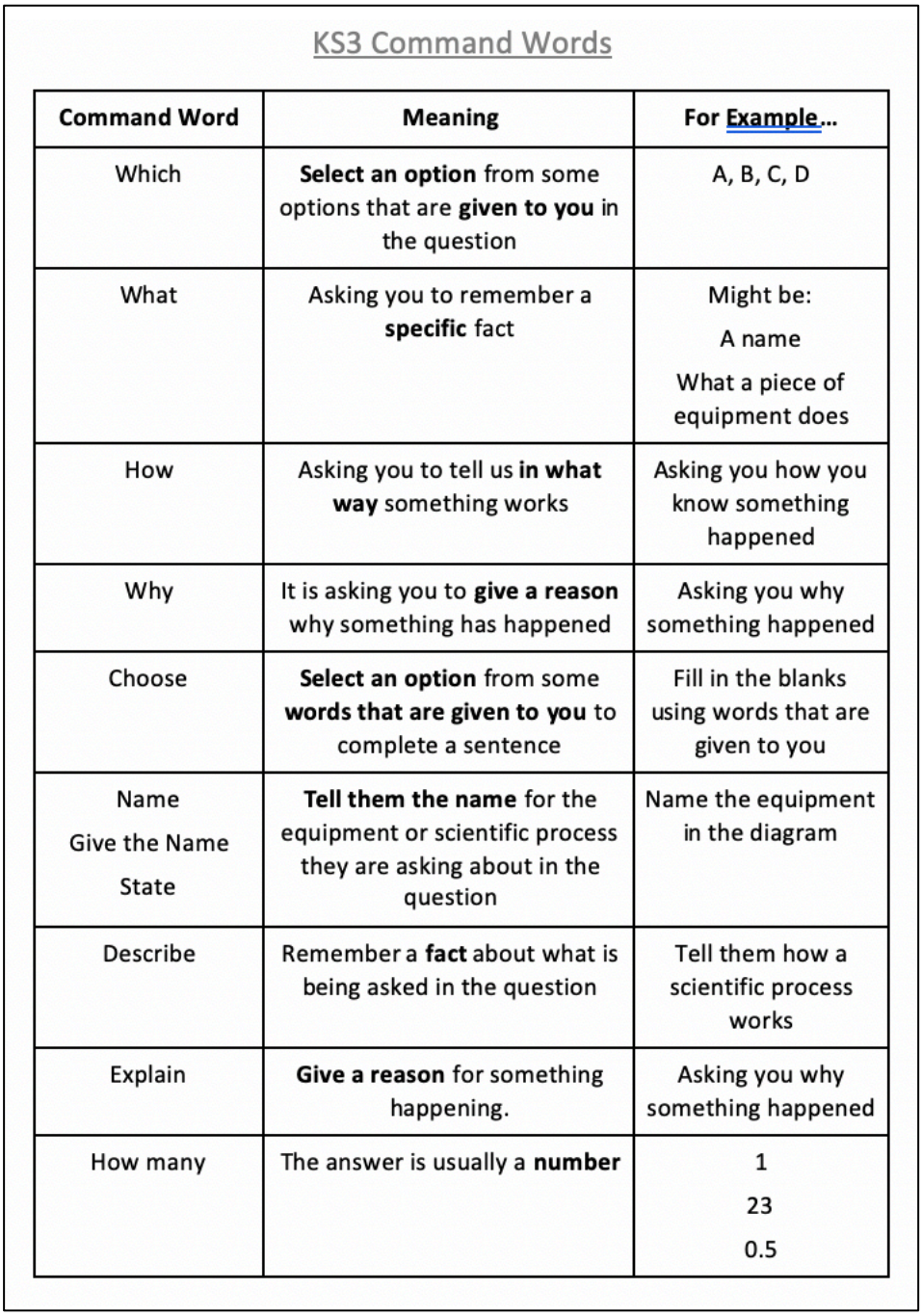
COMMAND WORDS (SCIENCE)

Command words are just the bits of a question that tell you what to do.

Command Word	What to do
Give/Name/State	Give a brief one or two word answer or a short sentence
Calculate	Use numbers given in the question to work out the answer.
Compare	Describe the **similarities and/or differences** between things, not just write about one.
Describe	Recall some facts, events or process in an accurate way.
Determine	Use **given data** or information to obtain and answer
Evaluate	Use the information supplied as well as your knowledge and understanding to **consider evidence for and against**.
Explain	Give **reasons** for something
Predict	Give a possible/logical outcome.
Show	Provide **evidence** to reach a conclusion.
Suggest	Use your **scientific knowledge** to work out what the answer might be
Outline	Write about the **main** points of a topic
Discuss	Write about a topic, considering different issues or ideas

Figure 1: *Key Stage 4 command word guidance sheet*

In order to design the command word guidance sheet, I again enlisted the help of Helen. We started off by exploring the command word guidance sheet given to GCSE students shown in ***Figure 1***. We both felt that we needed a more detailed sheet as Year 7s would be newly introduced to the concept of command words, and also felt examples may be helpful.



KS3 Command Words

Command Word	Meaning	For Example…
Which	**Select an option** from some options that are **given to you** in the question	A, B, C, D
What	Asking you to remember a **specific** fact	Might be: A name What a piece of equipment does
How	Asking you to tell us **in what way** something works	Asking you how you know something happened
Why	It is asking you to **give a reason** why something has happened	Asking you why something happened
Choose	**Select an option** from some **words that are given to you** to complete a sentence	Fill in the blanks using words that are given to you
Name Give the Name State	**Tell them the name** for the equipment or scientific process they are asking about in the question	Name the equipment in the diagram
Describe	Remember a **fact** about what is being asked in the question	Tell them how a scientific process works
Explain	**Give a reason** for something happening.	Asking you why something happened
How many	The answer is usually a **number**	1 23 0.5

Figure 2: *Part of Key stage 3 command word guidance sheet produced for the intervention.*

We then wondered if the command words in Key Stage 3 topic tests were similar to those found at Key Stage 4, so we looked through a sample of Year 7, 8, and 9 topic tests to pick out command words. We found that there were a wider range of command words used, so we then created a list of them and discussed definitions and examples that we felt the lowest attaining Year 7s would be able comprehend. Part of the guidance sheet is shown ***Figure 2*** with the full resource available to view in ***Appendix I***.

To design the 'walking talking mock' for the first test which would be the Particles topic for all classes, I wanted to emulate the PowerPoint presentation style of the Key Stage 4 equivalent. I also wanted to make sure that it was easy enough for teachers to use, without requiring too much additional input, meaning all students would receive a similar experience. The full 'walking talking mock' can be found in ***Appendix II***. I knew we needed to first explain what a command word was, as this would be unfamiliar to the Year 7s, as well as explaining how to highlight the command words in each question and use the command word guidance sheet to look up their meaning. I started with an example question, shown in ***Figure 3***, to model what a command word was, what it meant, and how they should look it up.

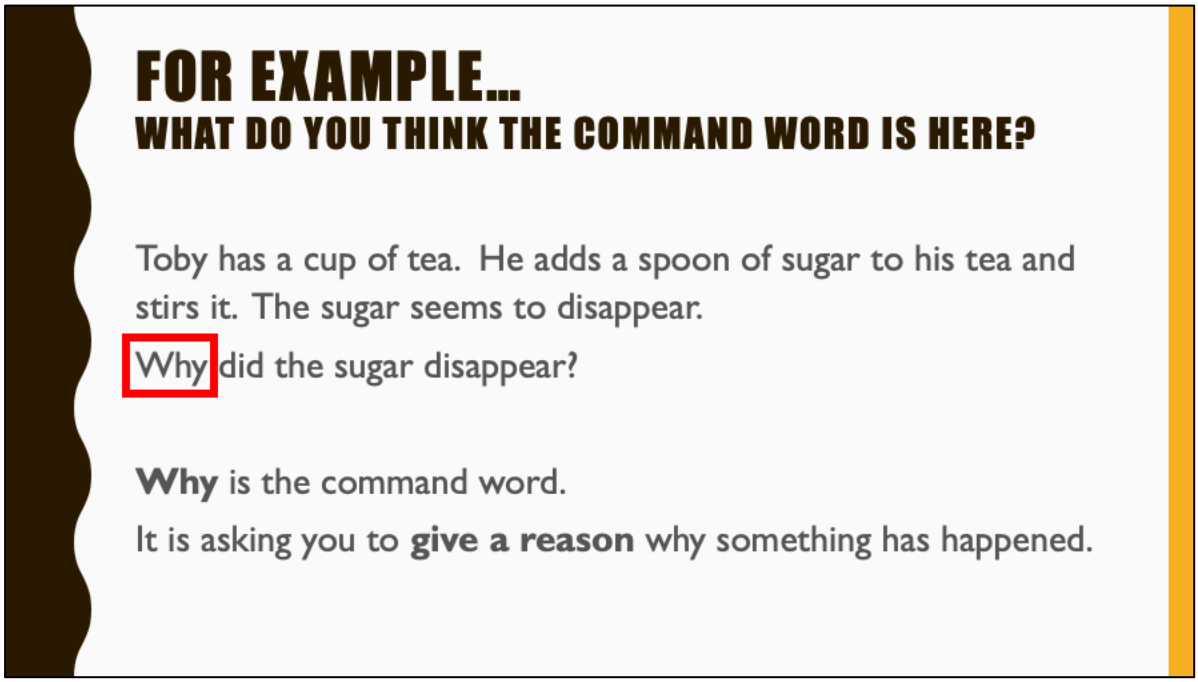


Figure 3: *Command word example for students at the start of the Particles topic 'walking talking mock'.*

I then went on to list some other exam top tips, as I was aware this was the first science assessment many of the students will have taken. The 'walking talking mock' then started. To begin with, each part of question 1 was read out. Students were given highly scaffolded support in order to identify the command word and look it up in their table. Students were then given a short amount of time to answer the question (as they were all 1-mark questions). Question 2 removed some of the scaffolding, simply getting students to identify the command word in each part of the question. Students were then given 1 minute per

mark to answer the sub-questions that made up question 2. I built in a timer into the presentation to make it as easy as possible for teachers to run the 'walking talking mock', as well as easy for students to see visually how much time they had left, with an audible alarm when the time ran out. Question 3 and question 4 were structured similarly to question 2, first getting students to identify the command words and then giving students 1 minute per mark to answer the questions. Finally, the last page on all KS3 tests is a topic feedback sheet. In addition to the topic feedback, in order to gauge student perceptions of the intervention, I added three simple questions for students to answer about the intervention. It was important these questions were not leading questions, and were quick and easy to answer, in order to take up minimal time, and to encourage a higher response rate. I showed the completed 'walking talking mock' to William, Drew and Helen and asked for their initial feedback. All found it to be well structured and easy to understand for both the teacher and the children.

Designing the second 'walking talking mocks' for the Energy & Electricity, and Reactive Science modules was easier as I used a similar model to the first 'walking talking mock'. They started off with a reminder about command words, but this time focused more on 'important information' in the question. 'Tick two boxes' or a question worth 2 marks would both be examples of important information in the question and highlighting this as well as command words further aimed to improve their exam technique further. Question 1 and 2 in both tests provided highly scaffolded guidance on how to break down each question, and the rest of the questions simply required students to highlight the command words and important information in each question, and then complete the question within a timed scenario (1 mark per minute). I was keen to make sure both 'walking talking mocks' had a similar structure and level of scaffolding to ensure students had a comparable experience when sitting different tests. The full 'walking talking mocks' for the Energy & Electricity, and Reactive Science topics can be found in ***Appendix III*** and ***Appendix IV*** respectively.

Data Collection and Analysis

With the purpose of investigating my first research question, I set out to determine what data I would need to collect in order to do so. The most logical data to use was end of

module test data, as it was standard across the year group, and collected at the end of each module as a matter of procedure. I aimed to compare the third and fourth tests, as well as the end of year test that students would have taken. This test mark data would have enabled me to compare the marks of the sample group with the marks of the control group. One drawback of using Year 7 test data was that each test only had a maximum of 25 marks, therefore the data had a relatively small range, meaning smaller or more subtle changes in attainment were more likely not to be shown after statistical analysis. In order to analyse the data, it was anonymised, and each child was given an ID number. I used SPSS to analyse the data, initially carrying out exploratory descriptive statistics on the data. I used Ennos' (2012) decision flowchart for statistical analysis shown in ***Figure 4*** to choose the most appropriate test: I was taking a measurement or rank; I was looking for differences between sets of measurements; I had two sets of measurements; and my two samples were not in matched pairs. This led me to either a parametric two-sample t-test (if the data was assumed to follow a normal distribution) or a non-parametric Mann-Whitney U test (if the data did not follow a normal distribution). When comparing each test of the sample group to the control group, I initially assumed normality as test data often follows a normal distribution and carried out two-sample t-tests to compare each test between the sample and control groups. I realised after carrying out a Kolgomorov-Smirnov test for normality that my data did not follow a normal distribution, and therefore carried out a Mann-Whitney U test.

With the view to exploring my second research question, in my authorisation letter to my headteacher, found in ***Appendix V***, I requested and was granted access to historical Year 7 science test data, from 2018 and 2017, as these years sat the same topic tests as the 2019 cohort. I also requested and was granted access to KS2, CATS and FFT data for all students in the three cohorts in order to be able to find comparable students at the high attaining, middle attaining and low attaining target ranges. My aim was to select 5 comparable students in each bracket of attainment from each cohort and see whether the sample group differed from the control group and previous cohorts. I chose only 5 students in each bracket because I aimed to look at students with the most similar previous data profiles, as

there would have been higher likelihood of picking up smaller changes after statistical analysis.

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at Ennos, A. R. (2012). Statistical and Data Handling Skills in Biology (3rd ed.). Pearson Education.

Figure 4: *Decision flowchart for statistical test* (Ennos, 2012)

After speaking to a maths teacher at my school for a second opinion, I devised a way to select my students, based on their FFT level, KS2 combined English and Maths SATs average score. I first filtered the results by FFT level. In Year 7, FFT targets in science range from 4 (high) to 1 (low). Therefore, I selected level 4 students for the high band, level 2/3 borderline students for the middle band and level 1 students for the low band. I then ranked the data according to their KS2 combined average SATs score. For the high band, I sorted it descending from high to low. For the low band, I sorted it ascending from low to high. Once this rank was in place, I sorted their average CATs score in the same way as the KS2 scores. This allowed me to find the median KS2 SATs and CATs result in the band, and then reduce the range around the median until I was left with 5 students around the median of the band. I discounted any students with missing data. The middle band proved more challenging. In the end, I decided to sort their KS2 SATs results ascending from low to high, finding the median value of their KS2 SATs and CATs score. I then chose results immediately surrounding the median, but as these were too numerous, I had to take the mode of the CATs scores within a ± 1.5 range around the median that gave me the smallest range of KS2 scores, which gave me 5 students. This process was repeated for each cohort (2019 sample, 2019 control, 2018 and 2017). The 2019 sample and control groups were then compared to the 2018 and 2017 using two-sample t-tests as I had assumed normality. Upon reflection, a Mann-Whitney U test may have been more appropriate as normality should not have been assumed.

Finally, to investigate teacher perceptions of the intervention, I decided I would carry out audio-recorded interviews with the three staff members teaching the sample groups in order to get a deeper insight into how they felt the interventions impacted the students, after the interventions had finished. I would then write a transcript of the interviews in order to analyse the answers to each question. To explore student perceptions of the intervention, I asked them to answer three short questions at the end of each test, which I then collated.

Ethical Considerations

A number of ethical considerations had to be made when designing this research project. This became clear whilst completing my Curec approval form and after reading the Ethical Guidelines for Educational Research report by the British Educational Research Association (British Educational Research Association, 2018).

Firstly, the year my project was to be completed was a difficult year in the science department as a number of staff were on long-term sick leave which meant that the many members of the science department were picking up the slack, so to speak, and this meant it was more crucial than ever to plan an intervention that required minimal increase to teacher workload. Had the circumstances surrounding the research project been different, I may have planned the intervention differently. I amended my final research question in order to reflect this. The only increase in workload for teachers was the short interview planned for a lunchtime or after school, and all teachers participating agreed to taking part knowing this would be a requirement.

Secondly, it was important to me that no student was disadvantaged by not receiving the intervention. In order to mitigate this, command word training and 'walking talking mocks' were planned to be introduced into the 2020 Year 8 scheme of work to make sure all classes had a similar experience. Selection of teachers to take part in the sample group was done without prejudice and was purely based on the experience of each individual and their place in the module rota.

Authorisation for the project was given by the headteacher of the school and all teachers involved signed a consent form informing them of what would be required of them and why. They were also informed of their right to withdraw from the study at any time. As I have no line management responsibilities in the school, there was no power imbalance that led to staff taking part, and no incentives were offered in exchange for taking part. Consent of students/parents was not required in the study as the research fell firmly under the school's modus operandi.

Outcomes of the research will be communicated back to the staff participants in a faculty meeting in the new school year.

All test data was immediately anonymised and assigned an ID number. The test data, along with signed consent forms and recordings of interviews was all stored on the school's internal Office365 OneDrive account. When reporting results, teachers were anonymised and 'fictionalised' in order to discuss their views about the intervention.

Coronavirus Impact and Other Challenges

Due to the coronavirus lockdown of schools in March 2020, my research project had to be adapted in a number of ways. Firstly, not all classes were able to complete their fourth topic test. This, coupled with the cancellation of end of year exams, meant that I only had viable data from the first three end of module tests. Therefore, I decided to place a greater weighting on the qualitative evidence provided by teachers and students, over the limited quantitative data analysis that I was able to achieve with the data collected. I also had to change the method of conducting the teacher interviews, from a face-to-face audio recorded interview, to a recorded online Zoom meeting (as became popular over lockdown to keep in touch with people locked down in their homes). These audio and video recordings were stored on the school's internal Office365 OneDrive system.

Another challenge I faced was the lack of student responses to the questions asked of them in test two. Myself and William's classes did not complete the three questions due to an oversight when creating the reactive science 'walking talking mock'. By the time the mistake had been noticed, too much time had lapsed to be able to retrospectively get their views.

Findings and Discussion

Research Question 1 - Quantitative Results

The first part of my investigation focused on assessing the impact of the intervention on the sample group, compared to the control group in each test. Initially I was only going to focus on tests 3, 4 and the end of year test, but due to the coronavirus lockdown of schools, I had to amend this to look at tests 1, 2 and 3.

Test 1

I compared the marks achieved by the 2019 sample group to the 2019 control group in test 1. As shown by the bar graph in ***Figure 5***, the mean mark of the sample group **(M = 18.63, SD = 4.62)** was only 0.4 higher than the mean mark in the control group **(M = 18.23, SD = 4.19)** and the standard error bars overlap significantly. A two-sample t-test showed that this difference was not significant t_{309} **= 0.79, p = 0.43**. These results show that there was no meaningful difference in the outcome of those who received the intervention vs. those who did not receive the intervention.

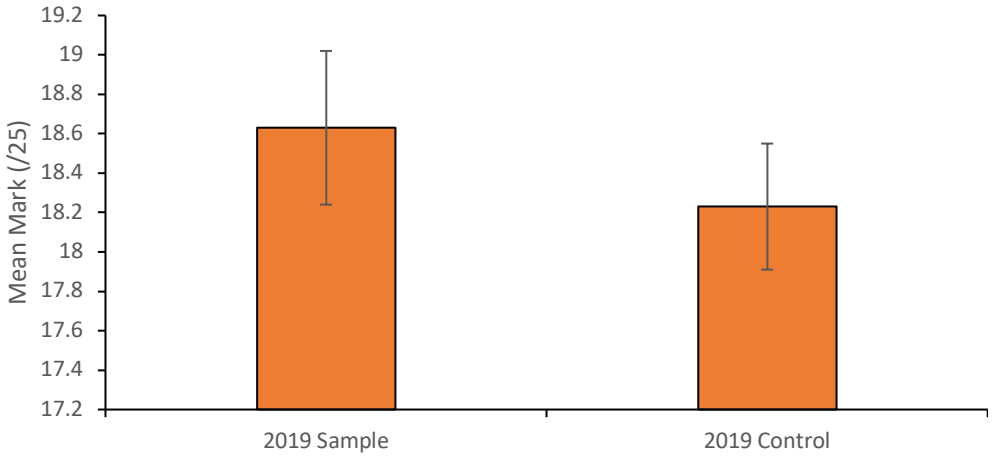


Figure 5: *Bar graph showing the mean (± standard error) of the mean marks of the sample and control groups in Test 1.*

I then did a Kolmogorov-Smirnov test for normality and it indicated that the 2019 sample group test 1 data does not follow a normal distribution, $D_{(140)}$ **= 0.125, p <0.05**. The same test on the 2019 control group reveals that it also does not follow a normal distribution, $D_{(171)}$ **= 0.116, p <0.05**. This means that the data has to be analysed with a non-parametric test. The non-parametric equivalent to the 2-sample t-test is the Mann-Whitney U test. As

shown in ***Figure 6***, the median for the 2019 sample group **(Mdn = 20.0)** was slightly higher than the median for the 2019 control group **(Mdn = 19.0)**, and the Mann-Whitney U test shows that there is no significant difference between the medians $U_{140, 171} = 11037.0, p = 0.236$. This means that when normality is not assumed, there is still no meaningful difference in the outcome of those who received the intervention vs. those who did not in Test 1. There are a number of reasons why this might be the case. As 25 marks is a relatively small number, subtle changes in attainment will not be noticeable. I also did not compare the overall prior attainment of the students in the sample group, vs. those in the control group. The particles test is also meant to be easier for students to grasp as we are aware it is the first science test that many students may have faced.

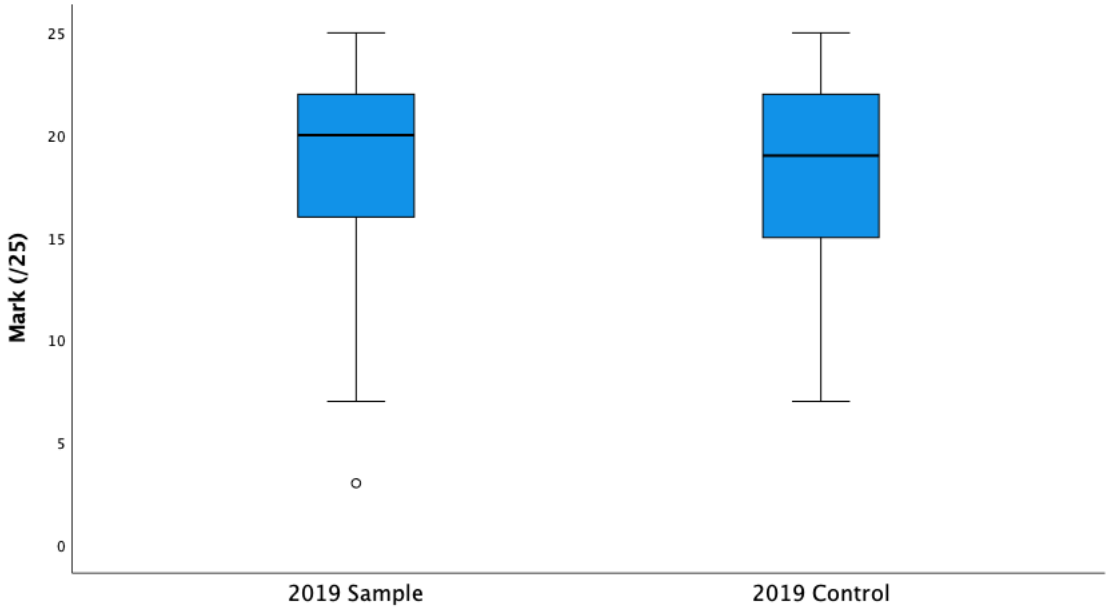


Figure 6: *Box and whisker plot showing the marks obtained by the sample and control groups in Test 1. Potential outliers have been represented with a circle.*

Test 2

I then compared the marks achieved by the 2019 sample group to the 2019 control group in test 2. As shown by the bar graph in ***Figure 7***, the mean mark of the sample group **(M = 16.7, SD = 3.91)** was only 0.4 higher than the mean mark in the control group **(M = 17.4, SD = 4.72)**. A Levene's test for equality of variances showed that the variance of the two samples is significantly different. A two-sample t-test showed that this difference was not

significant $t_{303.996}$ **= -1.385, p = 0.167**. These results show that there was no meaningful difference in the outcome of those who received the intervention vs. those who did not receive the intervention in Test 2.

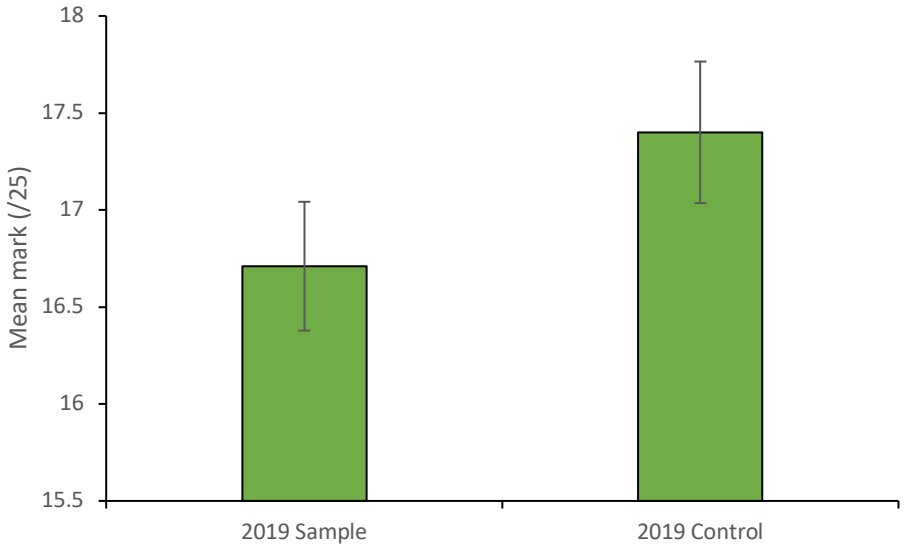


Figure 7: *Bar graph showing the mean (± standard error) of the mean marks of the sample and control groups in Test 2.*

A Kolmogorov-Smirnov test for normality indicated that the 2019 sample group test 2 data does not follow a normal distribution, $D_{(139)}$ **= 0.112, p <0.05**. The same test on the 2019 control group reveals that it also does not follow a normal distribution, $D_{(167)}$ **= 0.138, p <0.05**. This means that the data has to be analysed with a non-parametric test. As shown in ***Figure 8***, the median for the 2019 sample group **(Mdn = 17.0)** was slightly higher than the median for the 2019 control group **(Mdn = 18.0)**, and the Mann-Whitney U test shows that there is a significant difference between the medians $U_{139, 167}$ ***= 9926.0, p < 0.29***. This means that when normality is not assumed, the control group did significantly better than the sample group in test 2. This is unexpected but could be explained because a large proportion of the control group classes were doing a different test to the sample groups. The medical mysteries test may be easier than the energy & electricity and reactive science tests. In my professional opinion, I do find Year 7 find medical mysteries, which is all about the body, more relatable and have a better prior understanding of the topic due to personal experience than the other two topics which introduce lots of brand-new abstract concepts to them.

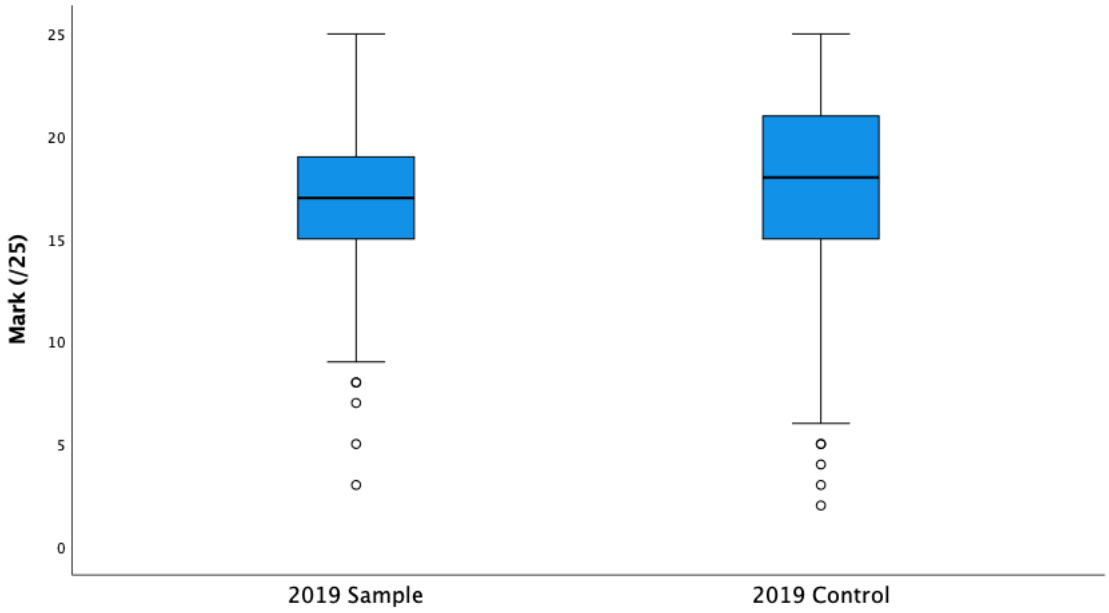


Figure 8: *Box and whisker plot showing the marks obtained by the sample and control groups in Test 2. Potential outliers have been represented with a circle.*

Test 3

Finally, I compared the marks achieved by the 2019 sample group to the 2019 control group in test 3. As shown by the bar graph in ***Figure 9***, the mean mark of the sample group **(M = 17.2, SD = 4.84)** was lower than the mean mark in the control group **(M = 17.8, SD = 3.71)**. A Levene's test for equality of variances showed that the variance of the two samples is significantly different. A two-sample t-test showed that this difference was not significant $t_{260.426}$ **= -1.223, p = 0.223**. These results show that there was no meaningful difference in the outcome of those who received the intervention vs. those who did not receive the intervention in Test 3.

A Kolmogorov-Smirnov test for normality indicated that the 2019 sample group test 3 data does not follow a normal distribution, $D_{(141)}$ **= 0.122, p <0.05**. The same test on the 2019 control group reveals that it also does not follow a normal distribution, $D_{(161)}$ **= 0.105, p <0.05**. This means that the data has to be analysed with a non-parametric test. As shown in ***Figure 10***, the median for the 2019 sample group **(Mdn = 18.0)** was the same as the median for the 2019 control group **(Mdn = 18.0)**, and the Mann-Whitney U test shows that there is

a significant difference between the medians $U_{141, 161} = 10856.50, p < 0.513$. This means that when normality is not assumed, the sample group had no significant difference to the control group in test 3.

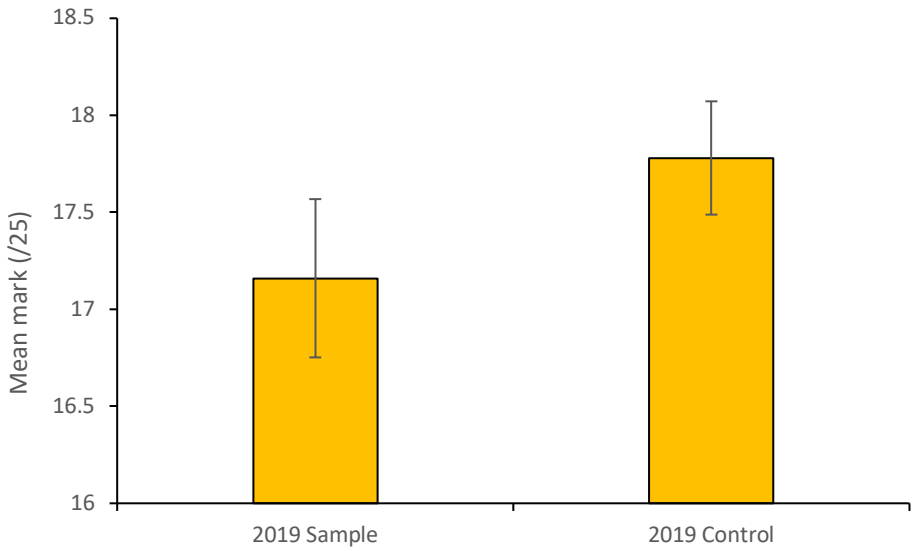


Figure 9: *Bar graph showing the mean (± standard error) of the mean marks of the sample and control groups in Test 3.*

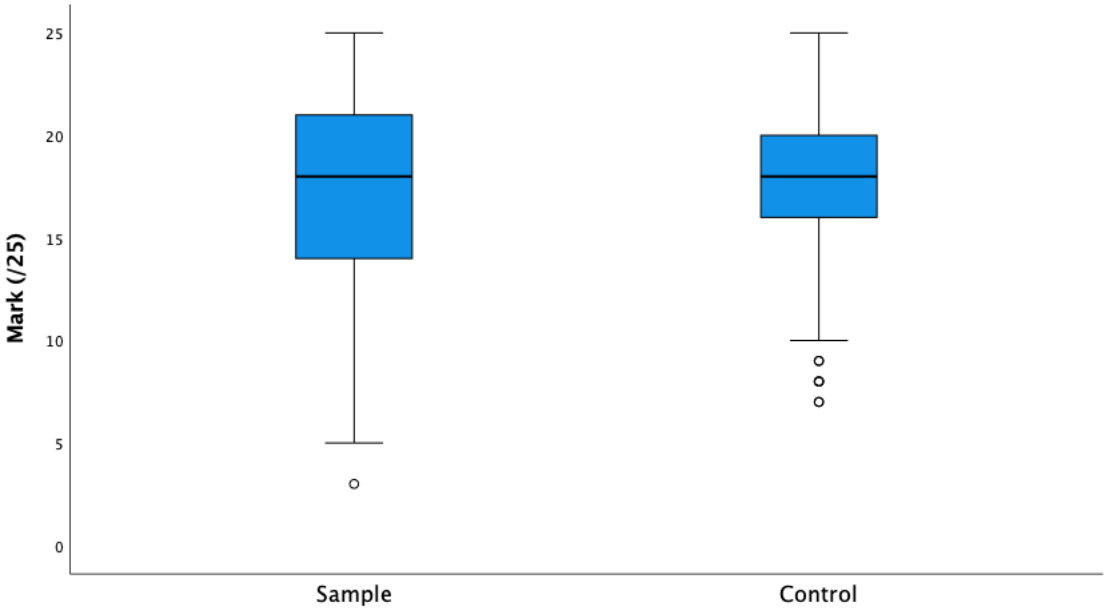


Figure 10: *Box and whisker plot showing the marks obtained by the sample and control groups in Test 3. Potential outliers have been represented with a circle.*

Research Question 2 - Quantitative Results

In order to investigate my second research question, I selected students from each cohort; 2019 sample, 2019 control, 2018 and 2017. 5 students from each cohort, with similar prior levels of attainment in KS2 SATs, CATs and FFT targets were chosen.

High Prior Attainment Band

Table 4: *Table showing the cohort means and standard deviations broken down by test in the high-attaining band.*

Test	Cohort	Mean	Std. Dev.
1	2019 S	23.6	1.673
1	2019 C	22.4	3.286
1	2018	21.6	3.782
1	2017	23.2	1.483
2	2019 S	21.0	2.236
2	2019 C	23.0	1.581
2	2018	19.2	3.564
2	2017	20.6	2.608
3	2019 S	22.6	0.548
3	2019 C	21.0	3.391
3	2018	20.6	3.912
3	2017	21.2	2.280

Table 4 shows the mean and standard deviation for each cohort, divided up by test. The results of the 5 students for the 2019 Sample cohort seem to have a very low standard deviation. Looking at ***Table 5***, none of the p-values are statistically significant as they are all greater than 0.05. This means that there was no statistical difference between the 2019

sample, 2019 control, 2018 and 2017 cohorts in terms of test results. Looking at **Figure 11**, you can see there is no meaningful difference in any of the bars and the error bars are all overlapping.

One weakness of this statistical analysis is that I assumed a normal distribution and have not tested for normality, and if that normality test came back as negative, then not completed a non-parametric test.

Table 5: Table showing the 2-sample t-test results for the high-attaining band. f and sig. only present if Levene's test for equality of variability came back as significantly varied.

Test	Cohort Comparison	t	df	p-value	f	Sig.
1	S - 2018	1.081	8	0.311	n/a	n/a
1	S - 2017	0.4	8	0.7	n/a	n/a
1	C - 2018	0.357	8	0.73	n/a	n/a
1	C - 2017	0.496	8	0.633	n/a	n/a
2	S - 2018	0.957	8	0.367	n/a	n/a
2	S - 2017	0.26	8	0.801	n/a	n/a
2	C - 2018	2.179	8	0.061	n/a	n/a
2	C - 2017	1.76	8	0.116	n/a	n/a
3	S - 2018	1.132	4	0.319	21.03	0.002
3	S - 2017	1.335	4	0.246	9.437	0.015
3	C - 2018	0.173	8	0.867	n/a	n/a
3	C - 2017	0.109	8	0.916	n/a	n/a

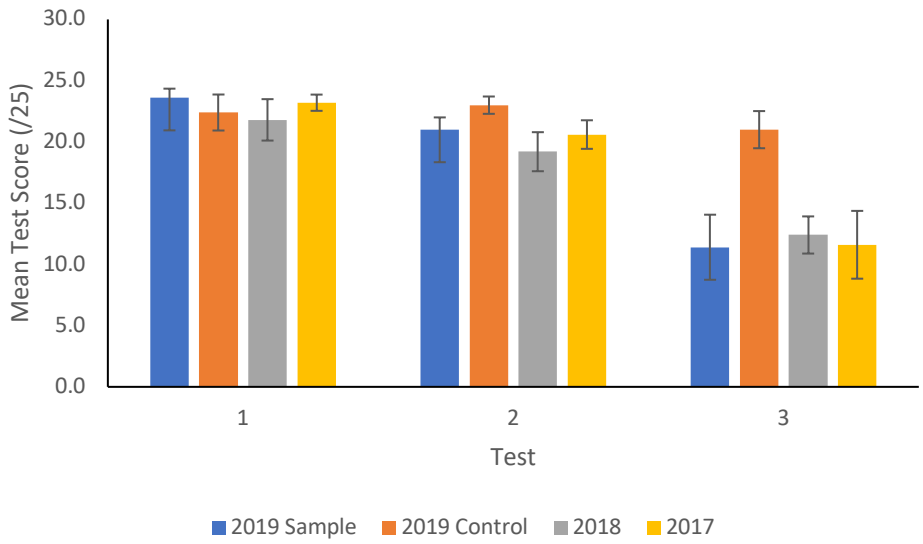


Figure 11 - Bar chart showing mean marks (± standard error) broken down by cohort and test in the high-attaining band.

Middle Prior Attainment Band

Table 6 shows the mean and standard deviation for each cohort, divided up by test. The standard deviations seem to be higher than the high attaining group. This is to be expected given the method students were selected by. Looking at ***Table 7***, two of the p-values are statistically significant as they are less than 0.05. This means that there was A statistically significant difference between the test 1 and 2 results when comparing the middle attaining 2019 control group to the middle attaining 2017 cohort. This could be down to chance or the sampling method. Looking at **Figure 12**, you can see there is only meaningful difference in the overlapping of the error bars in the results that were returned as significant.

Again, one weakness of this statistical analysis is that I assumed a normal distribution and have not tested for normality, and if that normality test came back as negative, then I have not completed a non-parametric test.

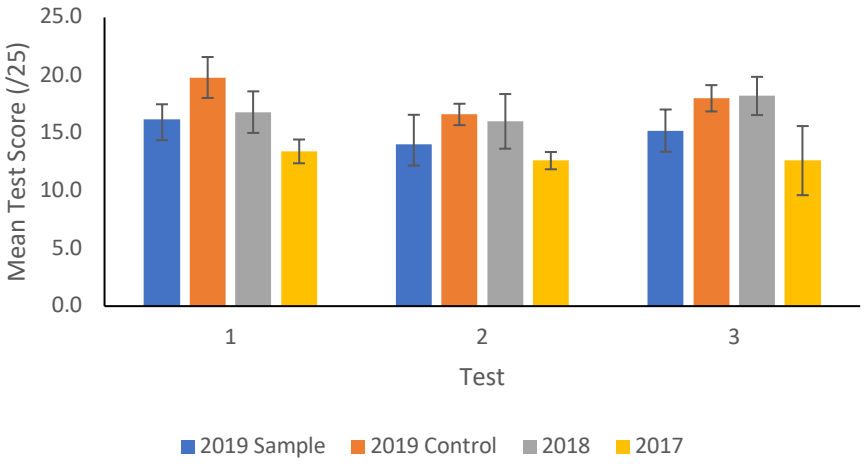


Figure 12: *Bar chart showing mean marks (±standard error) broken down by cohort and test in the middle-attaining band.*

Table 6: *Table showing the cohort means and standard deviations broken down by test in the middle-attaining band.*

Test	Cohort	Mean	Std. Dev.
1	2019 S	16.2	2.864
1	2019 C	19.8	3.962
1	2018	16.8	4.025
1	2017	13.4	2.302
2	2019 S	14.0	5.745
2	2019 C	16.6	2.074
2	2018	16.0	5.92
2	2017	12.6	1.673
3	2019 S	15.2	4.087
3	2019 C	18.0	2.55
3	2018	18.2	3.701
3	2017	12.6	6.693

*Table 7: Table showing the 2-sample t-test results for the middle-attaining band. f and sig. only present if Levene's test for equality of variability came back as significantly varied. *denotes a significant result where p<0.05.*

Test	Cohort Comparison	t	df	p-value	f	Sig.
1	S - 2018	0.272	8	0.793	n/a	n/a
1	S - 2017	1.704	8	0.127	n/a	n/a
1	C - 2018	1.188	8	0.269	n/a	n/a
1	C - 2017	3.123	8	**0.014***	n/a	n/a
2	S - 2018	0.573	8	0.583	n/a	n/a
2	S - 2017	0.523	8	0.615	n/a	n/a
2	C - 2018	0.236	5	0.822	7.954	0.022
2	C - 2017	3.357	8	**0.01***	n/a	n/a
3	S - 2018	1.217	8	0.258	n/a	n/a
3	S - 2017	0.741	8	0.48	n/a	n/a
3	C - 2018	0.1	8	0.923	n/a	n/a
3	C - 2017	1.686	8	0.13	n/a	n/a

Low Prior Attainment Band

Table 8 shows the mean and standard deviation for each cohort, divided up by test. Looking at ***Table 9***, none of the p-values are statistically significant as they are all greater than 0.05. This means that there was no statistical difference between the 2019 sample, 2019 control, 2018 and 2017 cohorts in terms of test results. Looking at **Figure 13**, you can see there is no meaningful difference in any of the bars and the error bars are all overlapping.

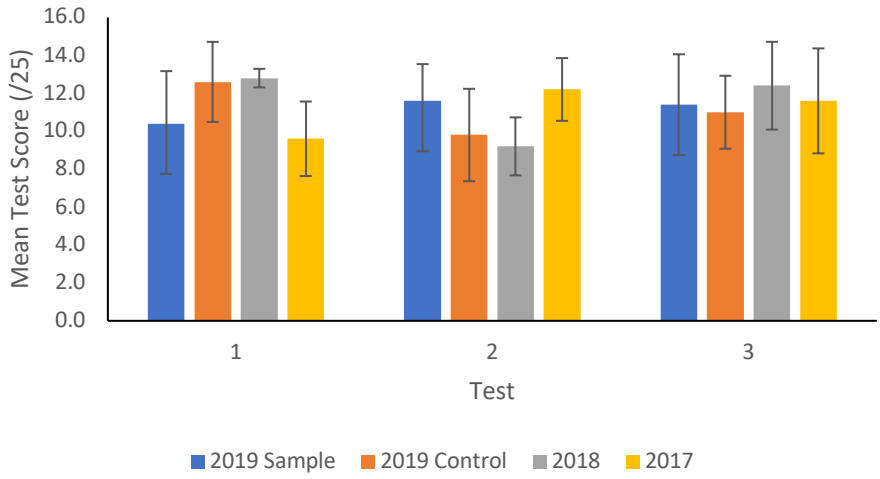


Figure 13: Bar chart showing mean marks (±standard error) broken down by cohort and test in the low-attaining band

Table 8: Table showing the cohort means and standard deviations broken down by test in the low-attaining band.

Test	Cohort	Mean	Std. Dev.
1	2019 S	10.4	6.189
1	2019 C	12.6	4.722
1	2018	12.8	1.095
1	2017	9.6	4.393
2	2019 S	11.6	4.336
2	2019 C	9.8	5.45
2	2018	9.2	3.421
2	2017	12.2	3.701
3	2019 S	11.4	5.941
3	2019 C	11.0	4.301
3	2018	12.4	5.177
3	2017	11.6	6.189

Table 9: Table showing the 2-sample t-test results for the low-attaining band. f and sig. only present if Levene's test for equality of variability came back as significantly varied.

Test	Cohort Comparison	t	df	p-value	f	Sig.
1	S - 2018	0.854	8	0.418	n/a	n/a
1	S - 2017	0.236	8	0.82	n/a	n/a
1	C - 2018	0.092	8	0.929	n/a	n/a
1	C - 2017	1.04	8	0.329	n/a	n/a
2	S - 2018	0.972	8	0.36	n/a	n/a
2	S - 2017	0.235	8	0.82	n/a	n/a
2	C - 2018	0.209	8	0.84	n/a	n/a
2	C - 2017	0.815	8	0.439	n/a	n/a
3	S - 2018	0.284	8	0.784	n/a	n/a
3	S - 2017	0.052	8	0.96	n/a	n/a
3	C - 2018	0.465	8	0.654	n/a	n/a
3	C - 2017	0.178	8	0.863	n/a	n/a

As with the high- and middle-attaining bands, a weakness of this statistical analysis is that I assumed a normal distribution and have not tested for normality, and if that normality test came back as negative, then I have not completed a non-parametric test.

Another weakness with the execution of the analysis of this data is that the sample selection was very narrow, at just 5 students per band in each cohort. Next time, considering increasing the sample size, to maybe just compare FFT levels as they already take into account KS2 SATs results.

Teacher Perceptions

Over the course of the entire research project, the part I found that I had personally gained the most from was the collaboration with William, Drew and Helen. Their insight, support and guidance has made me a better teacher. Due to the coronavirus lockdown, our interviews happened over Zoom and it was a welcome break from the mayhem of online teaching. My interview questions had not changed as a result of the lockdown, but I did use it as more of a talking point when discussing the intervention.

My first question to all three teachers was “Overall, what were your thoughts on the project aims?”. As we had all felt a bit overwhelmed by the demands of online teaching, I reminded all of the teachers what the project aims, outlined in their consent form, were. Right from the beginning, all three used words like “really valuable”, “really good” and “really useful” to describe the project. William commented that he felt the project made a “big difference to the students who traditionally struggle with tests to understand questions”. He then went on to comment that it may not have had the desired effect immediately, but that it will “benefit them later on when they hit more difficult questions”. Drew felt that the intervention was “really valuable because obviously we’ve successfully run walking talking mock GCSEs so I knew that they were a really effective intervention for kids and so I was really enthusiastic about trialling it in year 7”. Helen’s response reflected on our previous curriculum practice “I thought it was a really useful thing to find out because we do just tend to go into year 7 and say “here's a test” and expect them to know what to do with it but I know, obviously, they have done SATS in year 6, but it's a very different style of questioning we don't teach them in any way how to answer those questions.”. Helen went on to add “but also potentially really important for those students’ exam technique, and […] we’ve always looked at content and never really looked at the exam technique; we do leave it quite late really to sort of teach them how to answer the questions.”. Helen’s point aligned well with my initial motivation for investigating this particular subject.

My second question aimed to explore how they found the walking talking mock concept as a way of preparing students for exams. As the person that introduced the walking talking

mock concept to our school, Drew's explained that "given that with the first few tests they were transitioning from year 6 to year 7, it was a lot of modelling and a lot of repetition and a lot of just reassuring them, getting their feedback and then reassuring them about how it was going to run, outlining the benefits to them and it was reassuring, I think, for them because especially that first test in year 7, secondary school, they were like a rabbit caught in headlights, panic panic 'I've just done my SATs in primary school', and obviously to a greater or lesser extent, stressed about sitting those tests depending on the school and the pressure that was being put upon them by the school. So by and large they were pretty like 'oh my god I'm not quite sure about this'. When they recognise what we are trying to do with it and it was really just to support, to help them engage with the questions and some of the vocabulary and then how to structure their answers, the vast majority of them really bought into it and, I know they have that immediate feedback on the back of their test booklets, the ones that actually filled it in coherently and legibly, was really positive.". William agreed with Drew but went on further to comment about the test environment by saying it "made for better concentration and better classroom environment and better outcomes as well and I think it will actually stay as longer-term memory.". Helen compared the structure of the walking talking mock with her experiences at A Level; "I like it and I know how to do it; I've done it before. Possibly in not quite the same style as what we were doing with the year 7s obviously it's a lot more prescriptive with the PowerPoint and all the rest of it. With year 13 I've just literally given them a paper and said 'this question is worth 6 marks, you've got 6 minutes and then we'll go over it at the end', so I think it was slightly different in the fact that we introduced it a lot more and the way it was structured pulled away the tablecloth gradually as you went through". The common theme from all three answers were that the walking talking mock was well structured to guide year 7 students through the examination. Drew and Helen also confirmed that the investigation met one of my tests; to be familiar to teachers.

My next question was aimed at wanting to explore whether the teachers felt focusing on command words in the walking talking mock was the right approach to take as this differs slightly to the walking talking mocks we do at A Level. Drew gave an honest and insightful answer that I had not considered: "That's a good question because obviously command

words are something we really do focus on at GCSE and A Level to try and unlock what the question is asking them. So, I don't think it was useful, but I think it's useful more and more in later years when they understand what those command words mean and what that means in terms of the kind of answers they need to be giving. I think the brighter kids, whose vocabulary is good anyway, probably benefitted from it. The less-able kids perhaps who weren't quite so literate, I think it will probably prove its worth further down the line, I'm not sure with these tests because a lot of them were asking for reassurance during the test like 'Sir, can you explain this' or 'Sir, I don't understand', so, there was quite a bit of that going on from the lower ability kids.". William disagreed with the premise of Drew's answer and his reasoning was similar to mine when designing the intervention: "Yes, if we focus on the scientific discourse, we'd be giving them the answers and a lot of sense, during my revision I focus on scientific discourse, so I do keyword flashcards and that sort of stuff but we don't do exam flashcards so it was the right context you had to have had to already known the science, we weren't feeding them the answers we were pointing the answers in the right direction or how it's structured so yeah I think it was the right thing to do and I don't think we gave them an advantage over any other students in terms of knowledge but is giving the advantage of understanding the question which is the outcome basically.". Helen was undecided. She felt that on the one hand, command words do need to be taught and she also wasn't sure what else could have been focused on, but she wasn't sure how effective it was. "I wonder if maybe there's something we should introduce before we sit an exam that there's something that we try to use more in lessons because they sort of, I don't know, became almost dependant on those keywords like highlighting like mad because I'd encourage them to highlight them and maybe they less able ones losing the kind of the concept of what they were meant to be using those command words for.". In terms of implications for future practice, considering what else the walking talking mocks could focus on would be helpful given that all three teachers taking part had differing opinions.

I then wanted to probe as to whether they felt their students had made meaningful progress in their first two tests, compared to if they had sat them as normal in their opinion. All three teachers were optimistic but unsure whether their class would have done better compared to sitting the test as normal, but all three cited lower attaining students in their

answer, predicting them to have done better because of the time given to process the question, as well as having someone read the question out to them. Helen said “particularly ones with, you know, who would maybe normally have a reader in an exam and it obviously that's quite hard to manage if you don't have a TA as well so you felt like they were, you know, getting that support they needed and I agree I think again even those who maybe wouldn't have a reader in exam who but maybe have a you know lower literacy were able to access questions that they would have found too difficult, but I also think it makes them read the whole question […]”. The data analysis did not show this initially, but in the future a more in-depth analysis into each student to see if they improved over the course of the tests, rather than a cohort as a whole would be a relevant line of statistical enquiry.

Ideally, I would have wanted to get the data for test 4 as well as their end of year test, however this was not possible due to the coronavirus lockdown so I asked the teachers if they thought it would have had any impact in the future tests. Both Drew and Helen agreed that it would have been more difficult to see a trend in that data, whereas William felt they would have, even though it was a short project, as long as the reinforcement of the technique was there “Yeah, I think they would have, you know, I think they, I know it was a short project, they probably needed more and more reinforcement of the same message. I taught literacy for about 4 years, 5 years I think, in Australia and focus and activities every week whether it was the same activity, it did take them a lot of repetition before they actually understood the whole structure of things, they can understand the words and then putting it in the whole structure of things does take a lot of time. Literacy and understanding the questions did pay off and the school results with one year of literacy all the literacy, no, all the school’s actual literacy levels jumped up 2 years after 1 year of literacy just because they understood some things so I think it would have had a very good effect and it should have been more repetitive, should have done more tests, should have built it into the revision instead of the tests later on but, you know, I think there's a scope for more analysis there, more depth or just continuing with the project.”. William raised an interesting point that if more reinforcement of the techniques could have been done throughout the topics, there may have been a more noticeable result.

Next, I asked teachers how which types of students they felt would have benefitted the most from the intervention. All agreed that the higher attaining students would have been least affected by the intervention. Drew commented “for the less literate kids and by association, the less-abled kids, I think it was helpful from yes the decoding the question, but once they decode it, actually giving them the opportunity to actually access their knowledge and recall to write something down on paper.”. Both Helen and William gave similar answers.

My next question wanted to explore whether the teachers felt this intervention would be useful to roll out to other year groups, to encourage that repetition that William had mentioned. Drew felt that it could only improve performance and understanding of exam questions. **“**We’ve had a real push at GCSE and A Level so, for instance, in all the module tests at GCSE, and in all the A Level test, we get the command words on laminated cards now, and so having that familiarity beforehand is good so that they can look at a card and think ‘oh I know what “State” means compared to I know what “Explain” means’ and without having to stop and think about it and then the stress starts building up like ‘what kind of answer do I need to write here’. Hopefully, that won’t happen because they have got this experience that they’ve had in year 7 and maybe in year 8 and 9 to fall back on. So I think it can only help, genuinely, anything that helps with decoding questions, particularly with the content heavy GCSEs now and the time pressures and the fact that they are going to be sitting 6 separate GCSEs, for those kids it’s a massive stress, so anything that can help them, in terms of accessing the questions and therefore accessing the marks, it can only help.”. William agreed with Drew “Definitely, definitely I definitely see the same things people get wrong at GCSE so I spend a lot of time at GCSE post-test decoding the questions, decomposing the questions, like it’s an English lesson really, so we pull up all the questions, what is the questions asking, what does the question want in the answer, how do you know this.”. Helen agreed, and had already tried a similar intervention with her year 8 class. “Yeah I think it would, I definitely think at the start of the year, I mean, you know that I had a go at doing this with my year 8s, so I did, I think I did two tests like this or this sort of stuff, not quite such intervention as the ones that you'd written for year 7 so a bit more of a light touch but I think actually, you know, it is really good to remind them of those techniques because they're not going to remember them even from module to module let alone from

year to year so I think it's something that maybe in year 7 is really good to give them that really structured approach with lots of help and then in year 8 again beginning but with slightly less help so you're kind of you know gradually weaning them off you helping them but it just becomes the normal way of working because essentially if it's going to benefit them, it has to be just what they what they do without you having to remind them all the time and I definitely think it would be useful because I know we obviously use it at the end of year 11 but actually to use them through year 11 and year 10 and maybe to use it more as a teaching tool or in addition to using it as a you know an assessment tool because there's no reason why you can't give them in lessons you give them you know a long answer question to answer under exam conditions and then you go through it so you know what rather than just making it about the test, making it about learning exam technique which obviously we didn't do at lower school yeah but you could do it GCSE because we have so many exam questions in the lessons, so yeah I definitely think that it's- I don't think it has enough value to just do it at the beginning of year 7 and then it's almost telling the students that's not important anymore and then you have to sort of keep doing it to say actually yeah you know we did this in your second time we're going to look at it again and again because I mean the curriculum spirals so exam technique needs to be the same.".

My penultimate question focused on the teachers' views of the implications for our practice at school. Drew said "Yeah so again, I've spent, and other members of the department have spent, a lot of the sourcing exam questions to build in to the curriculum itself so every lesson now in theory will have a bank of exam questions because ultimately, this is the challenge for us, given that the curriculum is so content heavy, is that's fine one thing accumulating the knowledge, but as you know, trying to apply that in the context of an exam question, it's quite difficult to do that. A lot of my homework tends to be exam questions for that reason. Modelling exam questions is a really powerful teaching technique that I don't think we got enough time to do at GCSE. […] Again, if we can start making the links between test and, because it's no different to the types of questions they're going to get in their exams, then making that link is like a lightbulb going on in their heads 'Oh I can answer this "Explain" question', 'I can answer this "Discuss" question because I know what it means and I know what type of answer I need to frame on paper' so yeah, I think, you

know, you probably get the impression I'm a real keen advocate of this.". William also focused on the literacy of exam questions: "I think the implications are we take literacy a little bit more serious, I have a different view on literacy because I'm a little bit happy with the way you use literacy as punctuation, grammar, spelling, that's not literacy that's just good writing. Literacy is understanding words and communicating words, you can be as smart or as whatever but if you can't understand the words in front of you, that's being illiterate so it doesn't matter if they can't spell, doesn't matter if they can't punctuation and all that stuff because, well yeah it does matter, but I think the view of literacy is to comprehend, to decode, comprehend and communicate properly and that's literacy so I think the proper implications here are that we have to look at the communication and understanding of the exams separate to just science, just basic communicate, can you understand what I'm telling you.". Helen focused more on the feedback aspect of the intervention: "using it more as a teaching tool as well as an assessment tool, probably less relevant at lower school because we don't tend to really do a huge amount of exam questions and that the tests are kind of, they're quite short so you know and then use it using it as an assessment tool is valid but I think using it you know it year 10, 11 and sixth form it will be good for it to be written in somewhere into the scheme of work you know to give some options I mean in the past have used it with year 13 where I haven't had time to set a test and then mark it and feed it back to them although I would you know had limited time so I didn't feel that an hour of them sat there with me staring at them was good use of anybody's time, so actually you know them getting that immediate feedback. I did some work for one of my appraisals about multiple choice testing because obviously they get that at a level and the researchers found that unless they get the feedback the answers the actual answers within 24 hours of them taking multiple choice tests but they don't learn anything from it so that's a kind of like almost like you could start doing walking talking mock style multiple choice questions […]".

My final question asked the teachers whether they would change, remove or improve upon if we were to carry out the project again. Drew felt we could have incorporated the command word training in our test feedback: ". I think probably, personally, in terms of the feedback, maybe that 3rd test when you're handing the papers back, could do a bit more

work around that, around the command word strategies that we use, again to embed and kind of consolidate what we've been trying to do with the whole project. So maybe a focus on, you know, targeting feedback specific feedback, maybe going through the tests, looking at the common errors, looking at where students across the class had made errors and I don't think I did enough of really picking up on that and then applying that knowledge to the project and then doing something more enriching with it. Probably something in terms of feedback and the quality of the feedback based on what we've been trying to do with the strategy of the project.". Bill agreed that utilising feedback post-test was important, as well as increasing the number of year groups we offer the intervention to. Helen had a couple of improvements: "I think I probably would have liked to introduce them to the command words earlier rather than just leaving it to the exam.". She also felt giving more able students more flexibility to just get on with the test, so they were not sitting there bored and losing interest was important, however she recognised that it would have been difficult in a sample intervention for research setting. She also agreed with William that introducing it to more year groups was important. "I suppose, ideally, you'd want to make it a lighter and lighter touch each year, so it's really just a reminder of what they need to know, I mean possible even a sample question rather than the actual questions for the test so you know in a 40 minute year 10 test or year 11 test you could spend 10 minutes at the beginning going over a question so you could you know like you did for the year 7s, not beginning questions, but you could do that really structured approach for a question that's not in the test.".

I felt the teacher interviews gave great insight, and their experienced answers forced me to consider things I had not previously, like incorporating the training into timely test feedback.

Student Perceptions

I asked students three questions at the end of each test to gauge their perceptions towards the intervention. Sadly, external factors meant that not all classes had the opportunity to complete these questions for both tests, with significantly fewer students answering the questions at the end of test 2.

The first question asked to students was “Did you like the way we completed this test?”. I coded responses into two categories: Positive responses and Negative/Neutral Responses. ***Figure 14*** shows that responses were distinctly positive for both tests, which means students liked the walking talking mock approach. It is worth noting that these students will not have had any other end of module tests to compare to as these were the first two tests they sat.

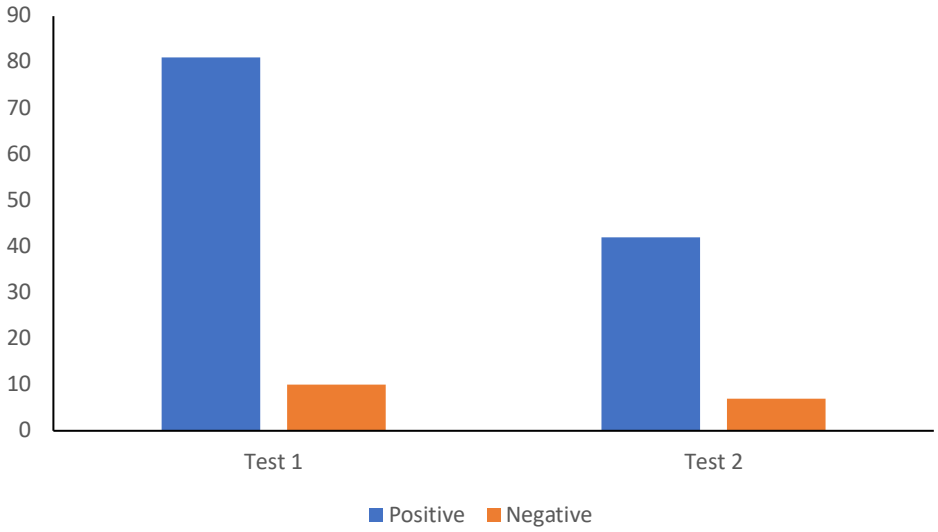


Figure 14: *A bar chart to show the responses to the question "Did you like the way we completed this test?" broken down by test.*

The second question asked to students was “What did you find helpful about the way we completed this test?”. Again, responses were coded into overarching themes, as shown in ***Figure 15***. Students found reading aloud and the increased time to answer questions most helpful. The ‘other’ responses ranged from commenting about the difficulty of the test to ‘I don’t know’. This corresponded to why the teachers felt it may have helped students.

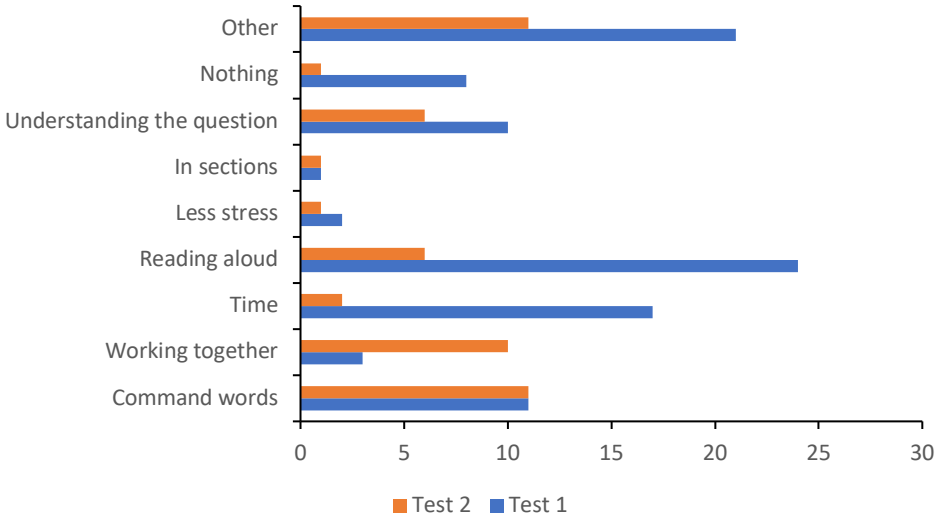


Figure 15: *Stacked bar chart to show the number of responses in each key theme in response to the question "What did you find helpful about the way we completed this test?", broken down by test.*

The third question simply asked the opposite of the second; "What did you find unhelpful about the way we completed this test?". Responses were coded and most responses showed that students found the walking talking mock approach to examination preferable. As the teachers predicted, the most common complaints were that students wanted to just get on with the test by themselves or found the time at one mark a minute too limiting/liberal.

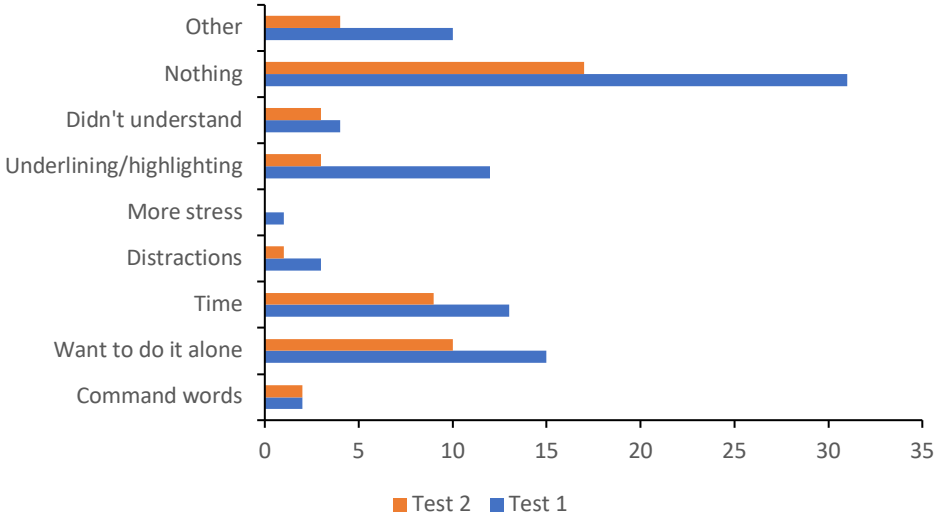


Figure 16: Stacked bar chart to show the number of responses in each key theme in response to the question "What did you find unhelpful about the way we completed this test?", broken down by test.

Conclusions and Implications

These quantitative results show that the intervention provided no significant benefit to the results of students who received it, nor did it give an advantage to any level of prior attaining student. However, there are a number of possible reasons for this. For instance, this may not provide immediate benefit, but repetition may have encouraged longer term benefits by the time they get to GCSE level. The coronavirus lockdown also prevented some data from being collected, and more pupil by pupil statistical analysis may have proved more descriptive than comparing whole cohorts. It is also worth noting that the sample size for the historical comparison was very small, so more descriptive results may have been achieved by increasing the sample size. If I were to repeat or extend the study, I would aim to improve and extend the command word training throughout the teaching of the module, as recommended by Childs et al. (2015). I would also increase the historical comparison to as many students with similar grade profiles as possible.

The teacher perceptions showed that, overall, the intervention was viewed favourably by teachers as they felt it would have benefitted students after more repetition, as well as increasing the confidence for lower-attaining students. They also agreed that rolling this approach out to all year groups to encourage familiarity throughout a student's school career would have benefitted the student in summative assessments.

The student perceptions were overwhelmingly positive to the approach. It has to be taken into account that students may not have a comparable test experience, as these were their first two science tests in their school career, however they still found it useful because of the time it allowed, and to have someone break down the question for them.

Going forward, I will present my findings to the science department in a faculty meeting, and if the Head of Department and the faculty agree that this approach to exam technique training is useful to implement from year 7, then we will write walking talking mocks for all modules during our summer curriculum development time when students are off sitting exams, ready for the following academic year. Hopefully, this would raise attainment in science not only in end of module tests, but also in summative external examinations in year 11 and year 13. However, the implications of this research are wider than just the science

department. I will be discussing my research at a TeachMeet event that the school organises, and presenting it to other faculties at the school. If other faculties think they could mirror or use some of the techniques learned through this study, then examination preparation from year 7 could be widely implemented across the school. There are also wider implications for Oxfordshire Partnership Schools, as sharing of this research through the Education Deanery could see other schools adopting a similar strategy, based on the teacher perceptions of the project.

Finally, I hope this project may start a discussion into researching the best methods to prepare students for examinations. As Madaus himself admitted, it is unavoidable for teachers to focus purely on developing independent skill in a subject when there is so much riding on the outcomes of students in their summative assessments (Madaus, 1998). Until widespread education reform occurs in England, we will continue to teach based on the outcomes of high-stakes assessment, and therefore educationalists should focus on how to support teachers to keep a healthy balance of fostering independent thinking in a subject and making sure students are prepared for their summative assessments.

References

Black, P., Daugherty, R., Ecclestone, K., Gardner, J., Harlen, W., James, M., Sebba, J., & Stobart, G. (2004). *A systematic review of research evidence of the impact on students, teachers and the curriculum, of the process of using assessment by teachers for summative purposes.* EPPI-Centre, University of London.

Booher-Jennings, J. (2008). Learning to label: socialisation, gender, and the hidden curriculum of high-stakes testing. *British Journal of Sociology of Education*, *29*(2), 149–160. https://doi.org/10.1080/01425690701837513

British Educational Research Association. (2018). *Ethical Guidelines for Educational Research*.

Cargill, F. (2019, December 16). How "walk and talk" mocks can boost GCSE results. *Times Educational Supplement*.

CFE Research. (2017). *'State of the nation' report of UK primary science education*.

Childs, P. E., Markic, S., & Ryan, M. C. (2015). The Role of Language in the Teaching and Learning of Chemistry. In *Chemistry Education* (pp. 421–446). https://doi.org/https://doi.org/10.1002/9783527679300.ch17

Class Teaching. (2014, November 27). *Walking Talking Exams*. Class Teaching. https://classteaching.wordpress.com/2014/11/27/walking-talking-exams/

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, *32*(1), 9–13.

Connelly, M. F., & Clandinin, J. D. (1990). Stories of Experience and Narrative Inquiry. *Educational Researcher*, *19*(5), 2–14.

Curtis, P. (2008, October 14). Sats for 14-year-olds are scrapped. *The Guardian*.

Curtis, P. (2009, May 6). Science Sats to be scrapped but maths and English tests expected to continue. *The Guardian*.

Dearing, R. (1994). *The National Curriculum and its Assessment: Final report*.

Department for Education. (2014). *The national curriculum in England*.

Dixson, D. D., & Worrell, F. C. (2016). Formative and Summative Assessment in the Classroom. *Theory Into Practice*, *55*(2), 153–159. https://doi.org/10.1080/00405841.2016.1148989

English, S. (2018, April 20). *Why I love…Walking Talking Mocks*. Susanenglish. https://susansenglish.wordpress.com/2018/04/20/why-i-love-walking-talking-mocks/

Ennos, A. R. (2012). *Statistical and Data Handling Skills in Biology* (3rd ed.). Pearson Education.

FFT. (n.d.). *The FFT Story*. Retrieved August 18, 2020, from https://fft.org.uk/the-fft-story/

Gersten, R., Baker, S., & Lloyd, J. W. (2000). Designing High-Quality Research in Special Education: Group Experimental Design. *Http://Dx.Doi.Org/10.1177/002246690003400101*, *34*(1), 2–18. https://doi.org/10.1177/002246690003400101

Gillard, D. (2018). *Education in England: a history*.

Green, S., & Oates, T. (2009). Considering alternatives to national assessment arrangements in England: Possibilities and opportunities. *Educational Research*, *51*(2), 229–245. https://doi.org/10.1080/00131880902891503

Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The

Impact of High-Stakes Testing on Teachers and Students in North Carolina. *Phi Delta Kappa International*, *81*(3), 199–203.

Kearns, L.-L. (2011). High-stakes Standardized Testing and Marginalized Youth: An Examination of the Impact on Those Who Fail. *Canadian Society for the Study of Education*, *34*(2), 112–130. https://doi.org/10.2307/canajeducrevucan.34.2.112

Madaus, G. (1998). The distortion of teaching and testing: High-stakes testing and instruction. *Peabody Journal of Education*, *65*(3), 29–46. https://doi.org/10.1080/01619568809538611

Madaus, G., & Clarke, M. (2001). *The adverse impact of high-stakes testing on minority students: Evidence from 100 years of test data*.

Minarechova, M. (2012). Negative impacts of high-stakes testing. *Journal of Pedagogy*, *3*(1), 82–100.

Moen, T. (2006). Reflections on the Narrative Research Approach. *International Journal of Qualitative Methods*, *5*(4), 56–69. https://doi.org/10.1177/160940690600500405786

Murray, J. (2013, January 14). "We are not cynical" say PiXL club schools. *The Guardian*.

Pintrich, P. R. (2000). The Role of Goal Orientation in Self-Regulated Learning. *Handbook of Self-Regulation*, 451–502. https://doi.org/10.1016/B978-012109890-2/50043-3

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and Self-Regulated Learning Components of Classroom Academic Performance. *Journal of Educational Psychology*, *82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond Cold Conceptual Change: The Role of Motivational Beliefs and Classroom Contextual Factors in the Process of Conceptual Change: *Http://Dx.Doi.Org/10.3102/00346543063002167*, *63*(2), 167–199. https://doi.org/10.3102/00346543063002167

PiXL. (n.d.). *The PiXL Club - About Us*. Pixl.Org.Uk. Retrieved August 29, 2020, from https://www.pixl.org.uk/about-us

Polesel, J., Dulfer, N., & Turnbull, M. (2012). *The Experience of Education: The impacts of high stakes testing on school students and their families*.

Popham, W. J. (2001). Teaching to the test. *Educational Leadership*, *58*(6), 16–20.

Pring, R., & Roberts, M. (2015). *A generation of radical educational change: Stories from the field*. Routledge.

Robbins, A. (2019, April 1). *Walking Talking Mocks: Not as useful as they seem?* Reflections on Education. https://adam-robbins.com/2019/04/01/walking-talking-mocks-not-as-useful-as-they-seem/

Scriven, M. (1981). *Evaluation Thesaurus* (3rd ed.). Edgepress.

Seuss, D., Prelutsky, J., & Smith, L. (1998). *Hooray for Diffendoofer Day!* Knopf.

Still, K. (2015, May 2). *Walking, talking mocks – worthwhile*. KristianStill. https://www.kristianstill.co.uk/wordpress/2015/05/02/walking-talking-mocks-worthwhile/#page-content

Taras, M. (2005). Assessment - Summative and Formative - Some Theoretical Reflections. *British Journal of Educational Studies*, *53*(4), 466–478. https://doi.org/10.1111/j.1467-8527.2005.00307.x

Thatcher, M. (1993). *The Downing Street Years*. HarperCollins London.

Education Reform Act, (1988) (testimony of UK Government).

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*, *28*(4), 499–509. https://doi.org/10.1016/S0346-251X(00)00035-X

Whetton, C. (2009). A brief history of a testing time: national curriculum assessment in England. *Educational Research*, *51*(2), 137–159. https://doi.org/10.1080/00131880902891222

Wilson, C. (2019, September 12). *Mini Walking Talking Mocks: what are they?* Pearson International Schools Blog. https://blog.pearsoninternationalschools.com/mini-walking-talking-mocks-what-are-they/

Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., Tomita, M. K., & Young, D. B. (2008). On the Impact of Formative Assessment on Student Motivation, Achievement, and Conceptual Change. *Applied Measurement in Education*, *21*(4), 335–359. https://doi.org/10.1080/08957340802347845

Zhao, M., Mu, B., & Lu, C. (2016). Teaching to the Test: Approaches to Teaching in Senior Secondary Schools in the Context of Curriculum Reform in China. *Creative Education*, *7*, 32–43. https://doi.org/10.4236/ce.2016.71004

Appendices

I. Command Word Sheet
II. Walking Talking Mock – Particles Topic
III. Walking Talking Mock – Energy and Electricity Topic
IV. Walking Talking Mock – Reactive Science Topic
V. Headteacher Authorisation
VI. William's Interview Transcript Sample
VII. Drew's Interview Transcript Sample
VIII. Helen's Interview Transcript Sample
IX. Student Response Results Sample

Appendix I – Command Word Guidance Sheet

KS3 Command Words

Command Word	Meaning	For Example…
Which	**Select an option** from some options that are **given to you** in the question	A, B, C, D
What	Asking you to remember a **specific** fact	Might be: A name What a piece of equipment does
How	Asking you to tell us **in what way** something works	Asking you how you know something happened
Why	It is asking you to **give a reason** why something has happened	Asking you why something happened
Choose	**Select an option** from some **words that are given to you** to complete a sentence	Fill in the blanks using words that are given to you
Name Give the Name State	**Tell them the name** for the equipment or scientific process they are asking about in the question	Name the equipment in the diagram
Describe	Remember a **fact** about what is being asked in the question	Tell them how a scientific process works
Explain	**Give a reason** for something happening.	Asking you why something happened
How many	The answer is usually a **number**	1 23 0.5

KS3 Command Words

Command Word	Meaning	For Example…
Plot/Draw	You need to **draw** something. You should use a pencil.	Graph Diagram
Complete	**Finish** a sentence/process/table that has been started for you by **filling in the blanks**.	Complete this table
Suggest	Use your **scientific knowledge** to work out what an answer **might be**.	Suggest why something happens.
Calculate	You will have to **do some maths** to work out the answer. **You can use a calculator** if you need to in Science.	Calculate the number of times something happens.
Estimate	**Using the information provided** in the question, give your **best guess**. The answer will usually be a number.	Estimate from the graph Estimate from the table
Tick [one] Give [two] ways	**Only** do as many things as the **question asks**. Doing **more or less** will make you **lose marks**.	Give three ways you can do something.
Look at… Use information/data from…	Include **data or information from the question** in your answer. This may be found in some text, a graph, or a table.	Using data from the graph in your answer…

Appendix II – Walking Talking Mock - Particles Topic

YEAR 7 PARTICLES TEST

WALKING TALKING MOCK

TO START…

We are going to start by thinking about the language of **exam questions**.

We need to look for **COMMAND WORDS**.

COMMAND WORDS are words in the question that tell us what to write about in the answer…

FOR EXAMPLE…
WHAT DO YOU THINK THE COMMAND WORD IS HERE?

Toby has a cup of tea. He adds a spoon of sugar to his tea and stirs it. The sugar seems to disappear.

Why did the sugar disappear?

FOR EXAMPLE…
WHAT DO YOU THINK THE COMMAND WORD IS HERE?

Toby has a cup of tea. He adds a spoon of sugar to his tea and stirs it. The sugar seems to disappear.

Why did the sugar disappear?

Why is the command word.

It is asking you to **give a reason** why something has happened.

SOME OTHER COMMON COMMAND WORDS

- **Which**
  - Select an option from some options that are given to you in the question (For example: A, B, C, D)
- **What**
  - Asking you to remember a specific fact (For example: a name, or what a piece of equipment does)
- **How**
  - Asking you to tell us in what way something works (For example: Asking you how you know something happened).
- **Choose**
  - Select an option from some words that are given to you to complete a sentence (For example: fill in the blanks).

SOME TOP TIPS

- When you first read a question, make sure you **underline** the command words in the question. Use the command words table on your desk to help you.
- Look at the **number of marks** for the question. That usually tells you how many things you need to do.
  - If there is one mark, then only tick one box, or give one fact.
  - Sometimes the question will tell you how many things to do.

  For example: Give TWO reasons…

  Only give TWO reasons or you risk losing marks if you give more or less than two reasons.

WE ARE NOW UNDER EXAM CONDITIONS

NO TALKING

QUESTION 1ai

Lets read the question out loud together and underline the command word.

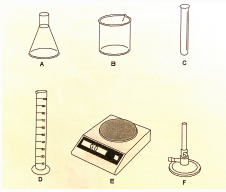
Q1ai

The diagram below shows six pieces of equipment.

a) Linda investigates how quickly sugar dissolves in water.

i) Which piece of equipment does she use to weigh 5g of sugar? Tick the correct box.

(1 mark)



Q1ai

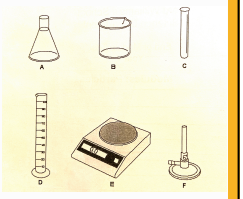
The diagram below shows six pieces of equipment.

a) Linda investigates how quickly sugar dissolves in water.

i) Which piece of equipment does she use to weigh 5g of sugar? Tick the correct box.

(1 mark)

WHICH is the command word. Use your table to find out what it means.



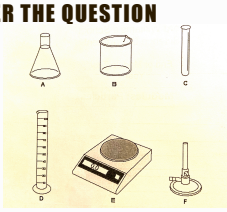
Q1ai – NOW ANSWER THE QUESTION

The diagram below shows six pieces of equipment.

a) Linda investigates how quickly sugar dissolves in water.

i) Which piece of equipment does she use to weigh 5g of sugar? Tick the correct box.

(1 mark)



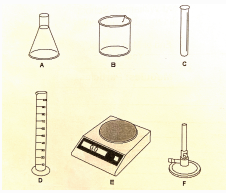
QUESTION 1aii

Lets read the next question out loud together and underline the command word.

Q1aii

ii) Which piece of equipment does she use to measure out 90cm³ of water? Tick the correct box.

(1 mark)

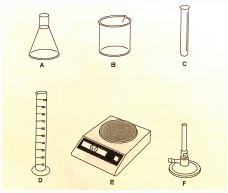


Q1aii

ii) Which piece of equipment does she use to measure out 90cm³ of water? Tick the correct box.

(1 mark)

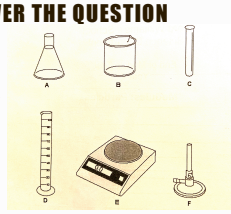
WHICH is the command word again. Use your table to remind yourself what it means.



Q1aii – NOW ANSWER THE QUESTION

ii) Which piece of equipment does she use to measure out 90cm³ of water? Tick the correct box.

(1 mark)



QUESTION 1bi

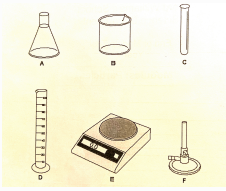
Lets read the next question out loud together and underline the command word.

Q1bi

b) Linda heats the water in a beaker.

i) Which piece of equipment shown is a beaker? Tick the correct box.

(1 mark)



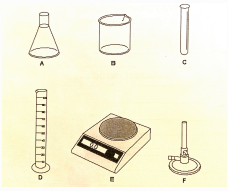
Q1bi

b) Linda heats the water in a beaker.

i) Which piece of equipment shown is a beaker? Tick the correct box.

(1 mark)

WHICH is the command word again. Use your table to remind yourself what it means.

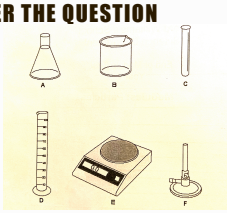


Q1bi – NOW ANSWER THE QUESTION

b) Linda heats the water in a beaker.

i) Which piece of equipment shown is a beaker? Tick the correct box.

(1 mark)



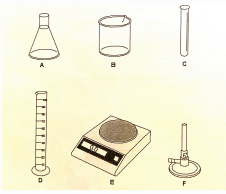
QUESTION 1bii

Lets read the next question out loud together and underline the command word.

Q1bi

ii) Which piece of equipment shown is used to heat water? Tick the correct box.

(1 mark)

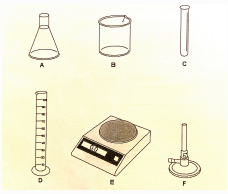


Q1bi

ii) Which piece of equipment shown is used to heat water? Tick the correct box.

(1 mark)

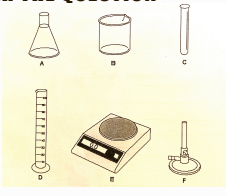
WHICH is the command word again. Use your table to remind yourself what it means.



Q1bi – NOW ANSWER THE QUESTION

ii) Which piece of equipment shown is used to heat water? Tick the correct box.

(1 mark)



QUESTION 1ci

Lets read the next question out loud together and underline the command word.

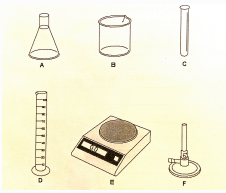
Q1bi

c) Linda adds 5g of sugar to the hot water.

i) She measures the time it takes for the sugar to dissolve. The equipment used for timing is **not** shown in the diagram.

What piece of equipment is used to measure the time taken?

(1 mark)



Q1bi

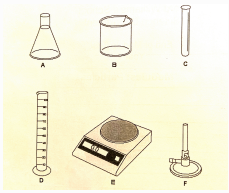
c) Linda adds 5g of sugar to the hot water.

i) She measures the time it takes for the sugar to dissolve. The equipment used for timing is **not** shown in the diagram.

What piece of equipment is used to measure the time taken?

(1 mark)

WHAT is the command word. Use your table to find out what it means.



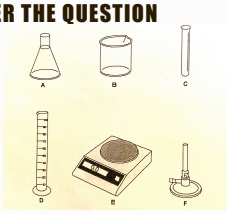
Q1bi – NOW ANSWER THE QUESTION

c) Linda adds 5g of sugar to the hot water.

i) She measures the time it takes for the sugar to dissolve. The equipment used for timing is **not** shown in the diagram.

What piece of equipment is used to measure the time taken?

(1 mark)



QUESTION 2

Lets read the next question out loud together.

Underline the command words in each question but do not answer the question.

Q2 COMMAND WORDS

a) WHY

bi) WHICH

bii) WHICH

c) WHICH

d) WHICH

ANSWER QUESTION 2.

WHEN YOU HAVE FINISHED, DO NOT MOVE ON.

YOU HAVE 5 MINUTES

QUESTION 3

Lets read the next question out loud together.

Underline the command words in each question but do not answer the question.

Q3 COMMAND WORDS

a) HOW

bi) WHICH

bii) WHICH

ci) WHY

cii) CHOOSE

ANSWER QUESTION 3.

WHEN YOU HAVE FINISHED, DO NOT MOVE ON.

YOU HAVE 8 MINUTES

QUESTION 4

Lets read the next question out loud together.

Underline the command words in each question but do not answer the question.

Q4 COMMAND WORDS

ai) WHAT

aii) WHAT

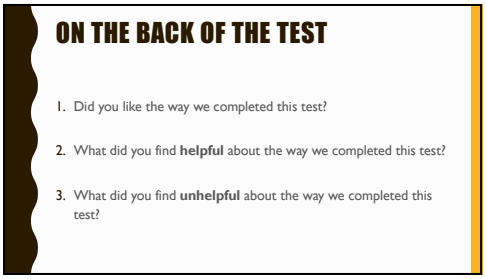
aii) GIVE THE NAME

bi) CHOOSE

ANSWER QUESTION 4.

IF YOU HAVE FINISHED, PLEASE FILL OUT THE TOPIC FEEDBACK SHEET

YOU HAVE 7 MINUTES


ON THE BACK OF THE TEST
1. Did you like the way we completed this test?
2. What did you find **helpful** about the way we completed this test?
3. What did you find **unhelpful** about the way we completed this test?

Appendix III – Walking Talking Mock – Energy and Electricity Topic

YEAR 7 ENERGY AND ELECTRICITY TEST

WALKING TALKING MOCK

TO START...

We are going to start by thinking about the language of exam questions.

We need to look for **COMMAND WORDS**.

Who can remember what a command word is?

COMMAND WORDS

- COMMAND WORDS are words in the question that tell us what to write about in the answer…

FOR EXAMPLE...
WHAT DO YOU THINK THE COMMAND WORD IS HERE?

Toby has a cup of tea. He adds a spoon of sugar to his tea and stirs it. The sugar seems to disappear.

Why did the sugar disappear?

FOR EXAMPLE...
WHAT DO YOU THINK THE COMMAND WORD IS HERE?

Toby has a cup of tea. He adds a spoon of sugar to his tea and stirs it. The sugar seems to disappear.

Why did the sugar disappear?

Why is the command word.

It is asking you to **give a reason** why something has happened.

SOME OTHER COMMON COMMAND WORDS

- **Which**
  - Select an option from some options that are given to you in the question (For example: A, B, C, D)
- **What**
  - Asking you to remember a specific fact (For example: a name, or what a piece of equipment does)
- **How**
  - Asking you to tell us in what way something works (For example: Asking you how you know something happened).
- **Choose**
  - Select an option from some words that are given to you to complete a sentence (For example: fill in the blanks).

SOME IMPORTANT INFORMATION

- Look at the **number of marks** for the question. That usually tells you how many things you need to do.
  - If there is one mark, then only tick one box, or give one fact.
  - Sometimes the question will tell you how many things to do.
    For example: Give TWO reasons…
    Only give TWO reasons or you risk losing marks if you give more or less than two reasons.
- When you first read a question, make sure you **underline** the command words in the question. Use the command words table on your desk to help you.

WE ARE NOW UNDER EXAM CONDITIONS

NO TALKING

QUESTION 1a

Lets read the question out loud together and underline the command word.

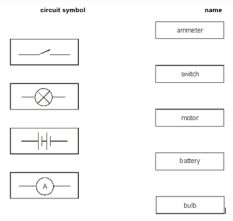
What is the important information in the question?

Q1a

Draw a line from each circuit symbol below to the correct name.

Draw only four lines.

(2 marks)



Q1a

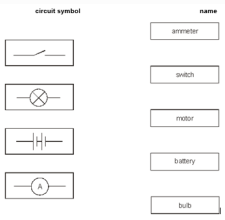
Draw a line from each circuit symbol below to the correct name.

Draw only four lines.

(2 marks)

DRAW is the command word. Use your table to find out what it means.

What is the important information in the question?



Q1a

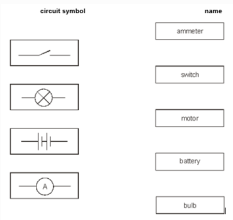
Draw a line from each circuit symbol below to the correct name.

Draw only four lines.

(3 marks)

DRAW is the command word. Use your table to find out what it means.

What is the important information in the question?

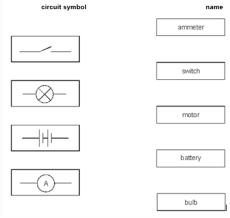


Q1a - NOW ANSWER THE QUESTION

Draw a line from each circuit symbol below to the correct name.

Draw only four lines.

(3 marks)



QUESTION 1b

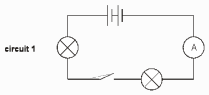
Lets read the next question out loud together and underline the command word.

Q1b

Fred made **circuit 1** as shown below.

Give the name of the part that is the energy source for the circuit.

(1 mark)



Q1b

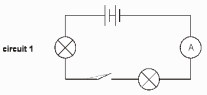
Fred made **circuit 1** as shown below.

Give the name of the part that is the energy source for the circuit.

(1 mark)

GIVE is the command word. Use your table to find out what it means.

What is the important information in the question?



Q1b

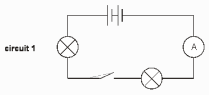
Fred made **circuit 1** as shown below.

Give the name of the part that is the energy source for the circuit.

(1 mark)

GIVE is the command word. Use your table to find out what it means.

What is the important information in the question?

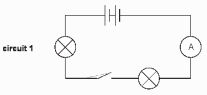


Q1b - NOW ANSWER THE QUESTION

Fred made **circuit 1** as shown below.

Give the name of the part that is the energy source for the circuit.

(1 mark)



QUESTION 1c

Lets read the next question out loud together and underline the command word.

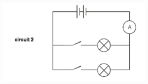
Q1c

Fred then made circuit 2 as shown below.

In the table below, tick a box to show whether **circuit 1** and **circuit 2** are series or parallel circuits.

Tick only **two** boxes.

(1 mark)



	series	parallel
circuit 1		
circuit 2		

Q1c

Fred then made circuit 2 as shown below.

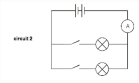
In the table below, tick a box to show whether **circuit 1** and **circuit 2** are series or parallel circuits.

Tick only **two** boxes.

(1 mark)

TICK is the command word. Use your table to find out what it means.

What is the important information in the question?



	series	parallel
circuit 1		
circuit 2		

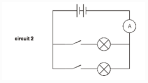
Q1c – NOW ANSWER THE QUESTION

Fred then made circuit 2 as shown below.

In the table below, tick a box to show whether **circuit 1** and **circuit 2** are series or parallel circuits.

Tick only **two** boxes.

(1 mark)



	series	parallel
circuit 1		
circuit 2		

QUESTION 2i

Lets read the next question out loud together and underline the command word.

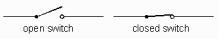
Q2i

A pupil makes some electrical circuits.

Each circuit contains a motor and a buzzer.

The motor and the buzzer can be switched on or off by three switches, A, B and C.

Symbols for an open switch and a closed switch are shown below.



By each circuit diagram, state whether:

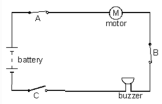
the motor is on or off;

the buzzer is on or off.

i) The motor is…

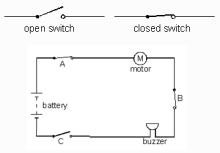
The buzzer is…

(1 mark)



Q2i

A pupil makes some electrical circuits.
Each circuit contains a motor and a buzzer.
The motor and the buzzer can be switched on or off by three switches, A, B and C.
Symbols for an open switch and a closed switch are shown below.



By each circuit diagram, state whether:
the motor is on or off;
the buzzer is on or off.

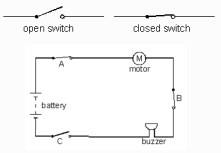
i) The motor is…
The buzzer is…
(1 mark)

STATE is the command word. Use your table to find out what it means.

What is the important information in the question?

Q2i

A pupil makes some electrical circuits.
Each circuit contains a motor and a buzzer.
The motor and the buzzer can be switched on or off by three switches, A, B and C.
Symbols for an open switch and a closed switch are shown below.



By each circuit diagram, state whether:
the motor is on or off;
the buzzer is on or off.

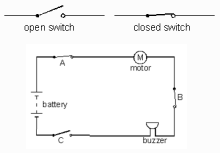
i) The motor is…
The buzzer is…
(1 mark)

STATE is the command word. Use your table to find out what it means.

What is the important information in the question?

Q2i – NOW ANSWER THE QUESTION

A pupil makes some electrical circuits.
Each circuit contains a motor and a buzzer.
The motor and the buzzer can be switched on or off by three switches, A, B and C.
Symbols for an open switch and a closed switch are shown below.



By each circuit diagram, state whether:
the motor is on or off;
the buzzer is on or off.

i) The motor is…
The buzzer is…
(1 mark)

ANSWER QUESTION 3.

WHEN YOU HAVE FINISHED, DO NOT MOVE ON.

YOU HAVE 4 MINUTES

QUESTION 3

Lets read the next question out loud together.

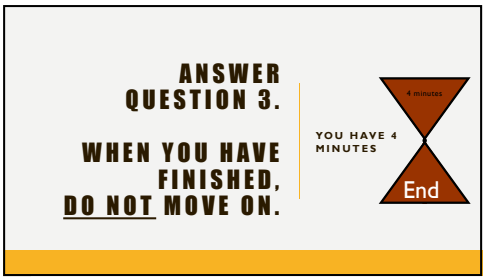
Underline the **command words** and the **important information to remember** in each question but do not answer the question.

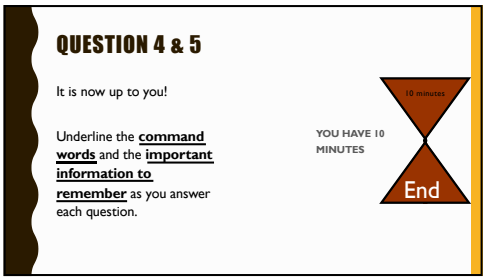
Q3 COMMAND WORDS & IMPORTANT INFORMATION

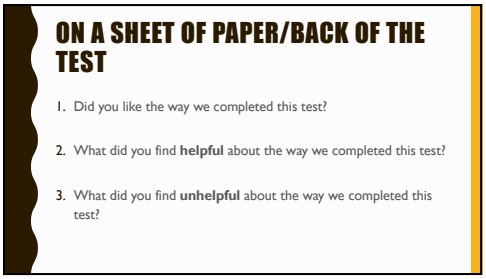
a) CHOOSE
BESIDE EACH PHOTOGRAPH WRITE THE NAME
2 MARKS

b) NAME
ONE
1 MARK

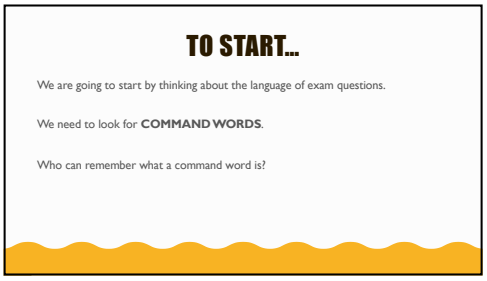
c) COMPLETE
PHOTOGRAPH (i)
1 MARK

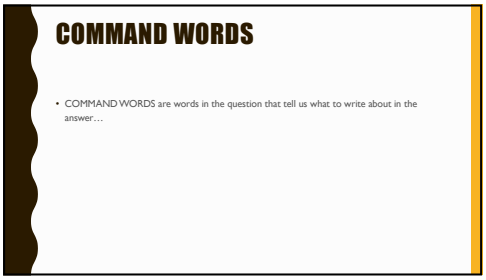

ANSWER QUESTION 3.
WHEN YOU HAVE FINISHED, DO NOT MOVE ON.
YOU HAVE 4 MINUTES
4 minutes
End

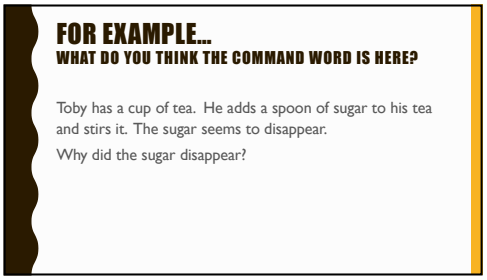

QUESTION 4 & 5
It is now up to you!
Underline the command words and the important information to remember as you answer each question.
YOU HAVE 10 MINUTES
10 minutes
End

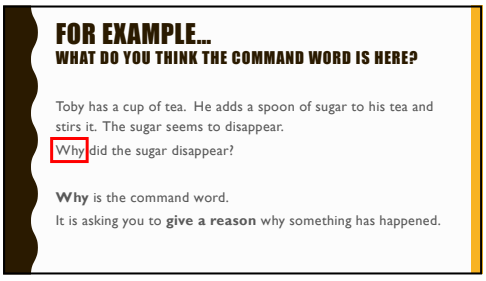

ON A SHEET OF PAPER/BACK OF THE TEST
1. Did you like the way we completed this test?
2. What did you find helpful about the way we completed this test?
3. What did you find unhelpful about the way we completed this test?

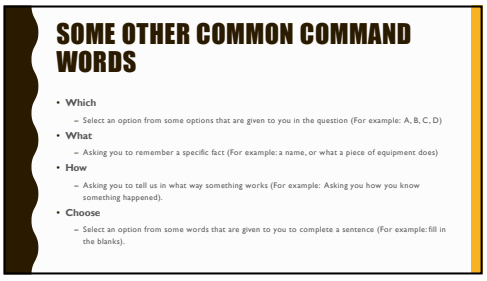
Appendix IV – Walking Talking Mock – Reactive Science Topic











SOME TOP TIPS

- When you first read a question, make sure you **underline** the command words in the question. Use the command words table on your desk to help you.
- Look at the **number of marks** for the question. That usually tells you how many things you need to do.
  - If there is one mark, then only tick one box, or give one fact.
  - Sometimes the question will tell you how many things to do.

  For example: Give TWO reasons…

  Only give TWO reasons or you risk losing marks if you give more or less than two reasons.

WE ARE NOW UNDER EXAM CONDITIONS

NO TALKING

QUESTION 1a

Lets read the question out loud together and underline the command word.

Q1a

Many chemicals are dangerous if not used carefully.

Read the **two** hazard descriptions. Look at the hazard labels.

Draw a line from each description to the correct label.

(2 marks)

hazard descriptions

POISONOUS
This will cause damage if swallowed

CORROSIVE
This can attack or dissolve many things

hazard labels


P


Q


R


S


T

Q1a

Many chemicals are dangerous if not used carefully.

Read the **two** hazard descriptions. Look at the hazard labels.

Draw a line from each description to the correct label.

(2 marks)

DRAW is the command word. Use your table to find out what it means.

hazard descriptions

POISONOUS
This will cause damage if swallowed

CORROSIVE
This can attack or dissolve many things

hazard labels


P


Q


R


S


T

Q1a

Many chemicals are dangerous if not used carefully.

Read the **two** hazard descriptions. Look at the hazard labels.

Draw a line from each description to the correct label.

(2 marks)

DRAW is the command word. Use your table to find out what it means.

hazard descriptions

POISONOUS
This will cause damage if swallowed

CORROSIVE
This can attack or dissolve many things

hazard labels


P


Q


R


S


T

Q1a - NOW ANSWER THE QUESTION

Many chemicals are dangerous if not used carefully.

Read the two hazard descriptions. Look at the hazard labels.

Draw a line from each description to the correct label.

(2 marks)

DRAW is the command word. Use your table to find out what it means.

hazard descriptions

POISONOUS
This will cause damage if swallowed

CORROSIVE
This can attack or dissolve many things

hazard labels

P Q R S T

QUESTION 1b

Lets read the next question out loud together and underline the command word.

Q1b

The label on a bottle of kitchen cleaner says:
It leaves kitchen and bathroom surfaces bright and shiny.
Another part of the label says:
Contains sulphamic acid
Irritating to eyes and skin
Keep out of reach of children

Which hazard label do you expect to see on the bottle?
Write the correct letter from the list above.

(1 mark)

hazard descriptions

POISONOUS
This will cause damage if swallowed

CORROSIVE
This can attack or dissolve many things

hazard labels

P Q R S T

Q1b

The label on a bottle of kitchen cleaner says:
It leaves kitchen and bathroom surfaces bright and shiny.
Another part of the label says:
Contains sulphamic acid
Irritating to eyes and skin
Keep out of reach of children

Which hazard label do you expect to see on the bottle?
Write the correct letter from the list above.

(1 mark)

WHICH is the command word again. Use your table to remind yourself what it means.

hazard descriptions

POISONOUS
This will cause damage if swallowed

CORROSIVE
This can attack or dissolve many things

hazard labels

P Q R S T

Q1b

The label on a bottle of kitchen cleaner says:
It leaves kitchen and bathroom surfaces bright and shiny.
Another part of the label says:
Contains sulphamic acid
Irritating to eyes and skin
Keep out of reach of children

Which hazard label do you expect to see on the bottle?
Write the correct letter from the list above.

(1 mark)

WHICH is the command word again. Use your table to remind yourself what it means.

hazard descriptions

POISONOUS
This will cause damage if swallowed

CORROSIVE
This can attack or dissolve many things

hazard labels

P Q R S T

Q1b – NOW ANSWER THE QUESTION

The label on a bottle of kitchen cleaner says:
It leaves kitchen and bathroom surfaces bright and shiny.
Another part of the label says:
Contains sulphamic acid
Irritating to eyes and skin
Keep out of reach of children

Which hazard label do you expect to see on the bottle?
Write the correct letter from the list above.

(1 mark)

WHICH is the command word again. Use your table to remind yourself what it means.

hazard descriptions

POISONOUS
This will cause damage if swallowed

CORROSIVE
This can attack or dissolve many things

hazard labels

P Q R S T

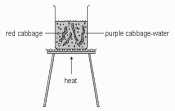
QUESTION 2

Lets read the next question out loud together and underline the command word.

Q2ai

Sharna boiled some red cabbage in water. The cabbage-water turned purple.

(a) Sharna mixed the purple cabbage-water with some other liquids.
She wrote the colours of the mixtures in a table as shown below.



Use the information in the table to answer parts (i) and (ii) below.

	colour of cabbage-water mixed with liquid	Is the liquid acidic, alkaline or neutral?
liquid 1	red	acidic
liquid 2	blue	alkaline
liquid 3	purple	neutral

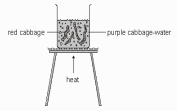
(i) Sharna mixed cabbage-water with colourless washing-up liquid.
The mixture turned **blue.**
What does this tell you about the washing-up liquid?

(1 mark)

Q2ai

Sharna boiled some red cabbage in water. The cabbage-water turned purple.

(a) Sharna mixed the purple cabbage-water with some other liquids.
She wrote the colours of the mixtures in a table as shown below.



Use the information in the table to answer parts (i) and (ii) below.

	colour of cabbage-water mixed with liquid	Is the liquid acidic, alkaline or neutral?
liquid 1	red	acidic
liquid 2	blue	alkaline
liquid 3	purple	neutral

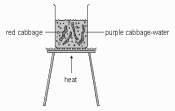
(i) Sharna mixed cabbage-water with colourless washing-up liquid.
The mixture turned **blue.**
What does this tell you about the washing-up liquid?
WHAT is the command word again. Use your table to remind yourself what it means.

(1 mark)

Q2ai

Sharna boiled some red cabbage in water. The cabbage-water turned purple.

(a) Sharna mixed the purple cabbage-water with some other liquids.
She wrote the colours of the mixtures in a table as shown below.



Use the information in the table to answer parts (i) and (ii) below.

	colour of cabbage-water mixed with liquid	Is the liquid acidic, alkaline or neutral?
liquid 1	red	acidic
liquid 2	blue	alkaline
liquid 3	purple	neutral

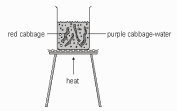
(i) Sharna mixed cabbage-water with colourless washing-up liquid.
The mixture turned **blue.**
What does this tell you about the washing-up liquid?
WHAT is the command word again. Use your table to remind yourself what it means.

(1 mark)

Q2ai – NOW ANSWER THE QUESTION

Sharna boiled some red cabbage in water. The cabbage-water turned purple.

(a) Sharna mixed the purple cabbage-water with some other liquids.
She wrote the colours of the mixtures in a table as shown below.



Use the information in the table to answer parts (i) and (ii) below.

	colour of cabbage-water mixed with liquid	Is the liquid acidic, alkaline or neutral?
liquid 1	red	acidic
liquid 2	blue	alkaline
liquid 3	purple	neutral

(i) Sharna mixed cabbage-water with colourless washing-up liquid.
The mixture turned **blue.**
What does this tell you about the washing-up liquid?
(1 mark)

WHAT is the command word again. Use your table to remind yourself what it means.

QUESTION 2aii

Lets read the next question out loud together and underline the command word.

Q2aii

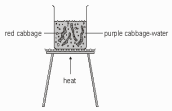
Sharna boiled some red cabbage in water. The cabbage-water turned purple.

(a) Sharna mixed the purple cabbage-water with some other liquids.
She wrote the colours of the mixtures in a table as shown below.

Use the information in the table to answer parts (i) and (ii) below.

(ii) Sharna then mixed cabbage-water with lemon juice.
Lemon juice is acidic.
What colour was the mixture?

(1 mark)



	colour of cabbage-water mixed with liquid	Is the liquid acidic, alkaline or neutral?
liquid 1	red	acidic
liquid 2	blue	alkaline
liquid 3	purple	neutral

Q2aii

Sharna boiled some red cabbage in water. The cabbage-water turned purple.

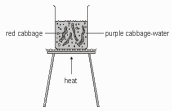
(a) Sharna mixed the purple cabbage-water with some other liquids.
She wrote the colours of the mixtures in a table as shown below.

Use the information in the table to answer parts (i) and (ii) below.

(ii) Sharna then mixed cabbage-water with lemon juice.
Lemon juice is acidic.
What colour was the mixture?

(1 mark)

WHAT is the command word again. Use your table to remind yourself what it means.



	colour of cabbage-water mixed with liquid	Is the liquid acidic, alkaline or neutral?
liquid 1	red	acidic
liquid 2	blue	alkaline
liquid 3	purple	neutral

Q2aii

Sharna boiled some red cabbage in water. The cabbage-water turned purple.

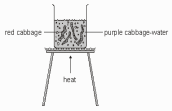
(a) Sharna mixed the purple cabbage-water with some other liquids.
She wrote the colours of the mixtures in a table as shown below.

Use the information in the table to answer parts (i) and (ii) below.

(ii) Sharna then mixed cabbage-water with lemon juice.
Lemon juice is acidic.
What colour was the mixture?

(1 mark)

WHAT is the command word again. Use your table to remind yourself what it means.



	colour of cabbage-water mixed with liquid	Is the liquid acidic, alkaline or neutral?
liquid 1	red	acidic
liquid 2	blue	alkaline
liquid 3	purple	neutral

Q2aii – NOW ANSWER THE QUESTION

Sharna boiled some red cabbage in water. The cabbage-water turned purple.

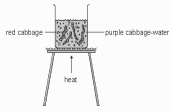
(a) Sharna mixed the purple cabbage-water with some other liquids.
She wrote the colours of the mixtures in a table as shown below.

Use the information in the table to answer parts (i) and (ii) below.

(ii) Sharna then mixed cabbage-water with lemon juice.
Lemon juice is acidic.
What colour was the mixture?

(1 mark)

WHAT is the command word again. Use your table to remind yourself what it means.



	colour of cabbage-water mixed with liquid	Is the liquid acidic, alkaline or neutral?
liquid 1	red	acidic
liquid 2	blue	alkaline
liquid 3	purple	neutral

QUESTION 2b

Lets read the next question out loud together and underline the command word.

Q2b

What is the name of a chemical which changes colour when it is mixed with acids or alkalis?

Tick the correct box.

(1 mark)

filtrate	☐	indicator	☐
non-metal	☐	solution	☐

Q2b

What is the name of a chemical which changes colour when it is mixed with acids or alkalis?

Tick the correct box.

(1 mark)

filtrate	☐	indicator	☐
non-metal	☐	solution	☐

WHAT is the command word. Use your table to find out what it means.

Q2b

What is the name of a chemical which changes colour when it is mixed with acids or alkalis?

Tick the correct box.

(1 mark)

filtrate	☐	indicator	☐
non-metal	☐	solution	☐

WHAT is the command word. Use your table to find out what it means.

Q2b – NOW ANSWER THE QUESTION

What is the name of a chemical which changes colour when it is mixed with acids or alkalis?

Tick the correct box.

(1 mark)

filtrate	☐	indicator	☐
non-metal	☐	solution	☐

WHAT is the command word. Use your table to find out what it means.

QUESTION 3

Lets read the next question out loud together.

Underline the **command words** and the **important information to remember** in each question but do not answer the question.

Q3 COMMAND WORDS & IMPORTANT INFORMATION

ai) WHICH
TICK THE CORRECT BOXES

aii) WHAT

b) WHAT

ci) WHY

cii) WHY

ciii) WHAT

ANSWER QUESTION 3.

WHEN YOU HAVE FINISHED, DO NOT MOVE ON.

YOU HAVE 6 MINUTES

QUESTION 4

Lets read the next question out loud together.

Underline the **command words** and the **important information to remember** in each question but do not answer the question.

Q3 COMMAND WORDS & IMPORTANT INFORMATION

a) WHAT
TICK TWO CORRECT BOXES
TWO MARKS

b) HOW

c) EXPLAIN
TWO MARKS

ANSWER QUESTION 4.

WHEN YOU HAVE FINISHED, DO NOT MOVE ON.

YOU HAVE 5 MINUTES

QUESTION 5 & 6

It is now up to you!

Underline the **command words** and the **important information to remember** as you answer each question.

YOU HAVE 10 MINUTES

Appendix V – Headteacher Authorisation Letter

Address removed to preserve anonymity

29th November 2019

Dear **[HEADTEACHER]**,

I am writing for permission to conduct a research project at **[SCHOOL]** as part of my MSc in Learning and Teaching at the University of Oxford. As a current Science teacher at the school, my project falls under the modus operandi of the school.

Project Details

Project Title: Formative use of Summative Assessment in Year 7 Science.

Project Supervisor: Dr Jane McNicholl, Associate Professor of Science Education, Department of Education, University of Oxford. Her contact email address is: jane.mcnicholl@education.ox.ac.uk.

Project Description: I aim to explore how using summative end-of-topic tests formatively can improve attainment in Year 7 Science. I aim to look at trends in attainment across the year group, historic Year 7 test data, general pupil and teacher perceptions of the interventions planned, as well as a more in-depth investigation into a small sample of students.

I propose to take five Year 7 classes, taught by four different teachers, and give these classes an make them complete the first two end-of-topic tests in a 'walking talking mock' style, as well as giving them more exam technique feedback during revision lessons. The remaining Year 7 classes will sit all end-of-topic tests as normal. I will then compare the data of the third and fourth end-of-topic tests across the year group, as well as compare key students across the year group who attained similarly in KS2, CATS and have the same FFT target levels in order to see if there was any significant difference in student outcomes. I will also be surveying students by asking 3 further questions within our usual end-of-topic feedback form and interviewing teachers on how they felt the different assessment style prepared students for future assessments. Students will complete this within their normal assessment pattern and so will not waste any learning time. Teachers will not have to prepare anything, as all resources will be made for them and given to them. They will give up one lunch time to complete the short interview mentioned above, and they have all agreed to this.

Benefits for the School

If this new way of working is found to be successful in improving student outcomes in Year 7, then embedding this into our science department scheme of work in every year is very simple with relatively little extra work, improving how our science department trains students for the exams they will face at the end of Year 11 and A-Level as soon as they start, encouraging good habits to form over their time at **[SCHOOL]**.

Data Storage

All data regarding the project (current and historic test data, audio recordings of staff interviews and transcripts of those interviews) will be anonymised immediately and then stored on the school's internal OneDrive for Business storage system. This data will then be deleted immediately after the assignment has been passed. Consent forms from Science staff members being recorded during an interview will be stored for three years on the school's Staff (N) drive to comply with research guidelines.

I require your signed consent below to confirm that this project falls under the school's modus operandi, and authorisation to access current and former Year 7 test data.

Please do not hesitate to contact me to discuss this further.

Kind regards,

[RESEARCHER]

CONSENT FORM

I, **[NAME]**, Headteacher of **[SCHOOL]** consent to **[RESEARCHER]** completing the research project above, entitled "Formative use of Summative Assessment in Year 7 Science". I confirm that this project falls under the school's modus operandi, and give authorisation to access current and former Year 7 test data.

Signed:

Print Name:

Date:

Consent removed to preserve anonymity.

Available on request.

Appendix VI – William's Interview Transcript Sample

Key - R: Researcher, T7: Teacher 7 (William)

R: So, this is being recorded, I have to tell you that of course your comments might be used in the dissertation project.

T7: Better lighting I'll see if I can get my lighting to be a bit better.

R: The video won't be used; I'll be making a transcript of this so don't worry.

T7: Okay fair enough.

R: So a lot of the questions are about the project itself so did you want me to read through the project aims just so you remind yourself of them?

T7: Sure, I can't remember them right now.

R: Yes, no problem. So, it was I aim to explore how using summative end of topic tests formatively can improve attainment in Year 7 science, I aimed to look at trends in attainment across the year group, historic year seven test data, general pupil and teacher perceptions of the interventions planned as well as a more in depth investigation into a small sample of students so I took five Year 7 classes taught by 4 different teachers, and gave those classes the first 2 end of topic tests in a walking talking mock style-

T7: I remember that, yeah.

R: -as well as giving them more exam technique feedback during revision lessons the remaining Year 7 classes sit an end of topic test as normal and then I was going to compare the 3rd and 4th end of topic tests across the year group as well as compare key students across the group who attained similarly at KS2 SATs and CATs and I have the same Fisher family trust target levels in order to see if there was any significant difference in student outcomes and then I surveyed students using 3 questions at the end of our end of topic feedback and I'm interviewing teachers on how they felt this different assessment styles prepared students for future assessments and students will complete, well had completed it, all within their normal assessment pattern and wouldn't waste any learning time.

T7: Okay

R: So the problem, of course, came with coronavirus and lockdown where we didn't get all the data we normally do so it was a lot more difficult to analyse but yeah so I suppose my first question is, overall, what were your thoughts on the project aims that I just sort of outlined?

T7: Yeah, no, I think it was really good. I think it was really good, I think vocabulary is a large part of understanding a test and so does the vocabulary from the discourse, so the vocabulary from the actual questions, what they mean, what the key words mean, that sort of stuff and I think it made a very big difference to the students who traditionally struggle with tests to understand the questions and that sort of stuff and it didn't make as big a difference to the students who naturally didn't see the issues. But it will benefit them later on when they hit more difficult questions because they can do it when it's easy but they get more difficult questions they really have to think about what the

question means so with students who found it easier to begin with being in Year 7, it's an easy year, the questions are fairly easy, it prepared them for how to think through different questions. So it was valuable for all of them, probably won't be apparent in the results later on for the higher achieving ones but for the lower achieving ones, the questions made more sense and they were able to answer them better and more fully as well. It was great, yeah.

R: Great so, how did you find the walking talking mock concept as a way of preparing students for exams?

T7: I thought it was really good you know, we kind of its there and they need that information so one of the things is that students' brains aren't as complicated as ours so they can learn something for the future and even we can't learn it for the future but when they have a need, an actual need there and then to understand the question and answer it, I think they engage better in it and they engage at a more learning level because they're actually recording it to use as their own answer they have that need they have made for better concentration and better classroom environment and better outcomes as well and I think it will actually stay as longer term memory.

R: Yep, I agree, did you think focusing on command words was the right approach to take during those walking talking mocks?

T7: Yes, if we focus on the scientific discourse, we'd be giving them the answers and a lot of sense, during my revision I focus on scientific discourse, so I do keyword flashcards and that sort of stuff but we don't do exam flashcards so it was the right context you had to have had to already known the science, we weren't feeding them the answers we were pointing the answers in the right direction or how it's structured so yeah I think it was the right thing to do and I don't think we gave them an advantage over any other students in terms of knowledge but is giving the advantage of understanding the question which is the outcome basically.

****Full transcript available upon request****

Appendix VII – Drew's Interview Transcript Sample

Key - R: Researcher, T1: Teacher 1 (Drew)

R: I obviously have to tell you this is being recorded and I might use quotes from you in the project

T1: Yeah

R: So my first question is – I've sent you on the consent form all the project aims

T1: Yes

R: I can read them to you if you'd like

T1: I've just set it to print, shall I go and grab it?

R: Yes, go grab a copy

T1: Right yes, so go on sorry you were saying

R: So it was overall, what were your thoughts on the project aims in general? I'll let you read

T1: Yep well really, I mean from my point of view, really valuable because obviously we've successfully run walking talking mock GCSEs so I knew that they were a really effective intervention for kids and so I was really enthusiastic about trialling it in Year 7 and I was going to talk to you actually if you thought maybe the first test of Year 8 and 9 perhaps should be walking talking mocks to just remind them of what they have done in Year 7. I don't know but maybe that's something for later but yeah, really really enthusiastic and engaged with it because I know how effective it can be. Absolutely bought into it, 100%.

R: Great! I know you run the walking talking mocks for Year 10 but how did you find that concept as a way of preparing students for exams in Year 7?

T1: Yeah I mean any technique, especially with Year 7, and given that with the first few tests they were transitioning from year 6 to Year 7, it was a lot of modelling and a lot of repetition and a lot of just reassuring them, getting their feedback and then reassuring them about how it was going to run, outlining the benefits to them and it was reassuring, I think, for them because especially that first test in Year 7, secondary school, they were like a rabbit caught in headlights, panic panic 'I've just done my SATs in primary school', and obviously to a greater or lesser extent, stressed about sitting those tests depending on the school and the pressure that was being put upon them by the school. So by and large they were pretty like' oh my god I'm not quite sure about this'. When they recognise what we are trying to do with it and it was really just to support, to help them engage with the questions and some of the vocabulary and then how to structure their answers, again the vast majority of them really bought into it and, I know they have that immediate feedback on the back of their test booklets, the ones that actually filled it in coherently and legibly, was really positive. Obviously, you might have looked at the data, but I think as they progressed onto doing it under normal conditions, I think they were reassured. I don't know what the data shows in terms of progress but certainly in terms of just settling them down, reassuring them, not being instinctively afraid of the questions, I think really helped from that point of view so they went into it with a much more positive attitude than they would have done without the walking talking practice beforehand, I think. That's my overall view, I think.

R: Okay great! I'd agree with you definitely, but did you think focusing on command words was the right approach to take in the walking talking mocks?

T1: You see, that's a good question because obviously command words are something we really do focus on at GCSE and A Level to try and unlock what the question is asking them. So, I don't think it was useful, but I think it's useful more and more in later years when they understand what those command words mean and what that means in terms of the kind of answers they need to be giving. I think the brighter kids, whose vocabulary is good anyway, probably benefitted from it. The less-abled kids perhaps who weren't quite so literate, I think it will probably prove its worth further down the line, I'm not sure with these tests because a lot of them were asking for reassurance during the test like 'Sir, can you explain this' or 'Sir, I don't understand', so, there was quite a bit of that going on from the lower ability kids. So it will be interesting, as I said, this time around Year 7, the brighter kids, great, I think the less-abled kids, probably further down the line you'll see its worth and effectiveness then I think.

R: Yeah, I definitely agree with that as well. Do you think that in the first 2 tests, because I haven't started crunching the data yet, do you think it would have improved pupil outcomes in the first 2 tests compared to those who didn't have a walking talking mock style of exam in the first 2 tests?

T1: I think it can only have benefit – was it the benefit from the walking talking mock you mean, in terms of outcomes?

R: Yes

T1: They certainly wouldn't have done any worse, I can only think it would have benefitted. I know some of the students expected, particularly the more difficult questions, not to have engaged with them at all, not to have written anything at all, at least they were making an attempt, it might be a wrong attempt but at least they – again it comes back to that confidence thing, I think because they understand what some of those difficult command words meant it meant that they had a fighting chance of putting some sort of coherent answer together, whether it was right, I don't know, and that is the thing we find with GCSE extended questions is a lot of the time, kids will come up with 0 or 1 mark because they have that 'oh my god, how am I going to answer this question', without actually just unpicking the question and then unpacking it and then coming up with some answer at least which allows them to get at least some marks. I think that's a huge benefit of this, is the fact that, again, it's about building confidence and decoding questions. I think that will be, and probably was for some kids in these tests, but certainly down the line, a really powerful impact of doing the walking talking mocks.

R: Yeah, and that's sort of my next question, do you think it would have impacted pupils in following tests because we didn't really get too much of that data because of the coronavirus lockdown.

T1: Yeah

****Full transcript available upon request****

Appendix VIII – Helen's Interview Transcript Sample

Key - R: Researcher (XAn), T3: Teacher 2 (Helen)

R: I of course have to let you know that it's recorded

T3: Yeah

R: Bits of it will be like quoted in the project and stuff like that but no, thank you so much. So, the first question is, overall, what were your thoughts on the project aims? So, if you want me to read you the project aims, I'm more than happy to.

T3: Yeah if you could just remind me, that would be great thank you

R: No problem, so I was exploring how using summative end of topic tests formatively throughout the year can improve attainment in Year 7 science, so looking at trends in attainment across the year group, looking at, sort of, Year 7 historic data to see if it changed and teacher perceptions and pupil perceptions of those interventions that we ran, which were the walking talking mocks and all the command word training as well as a bit more in depth investigation into a small sample of comparable students and so we took 5 Year 7 classes, taught by 4 different teachers, giving them, sort of, giving them the first 2 end of topic tests in a walking talking mock style as well as giving them more, sort of, exam technique feedback during revision lessons and all that command word training and the remaining Year 7 classes all sat the end of topic tests as normal and then comparing the 3rd and 4th end of topic tests as well as the end of year test and comparing key data – well key students that were comparable across the year group, so ones that had the same KS2 scores, CAT scores, additional family trust targets etc. To see if there was a difference in student outcomes. Now of course the COVID lockdown has thrown all that out the window

T3: Scuppered everything!

R: Yeah so we didn't get much of that data! But that's okay, I just have to write about other things so I, sort of, increased the number of questions I was going to ask but it shouldn't take too long. That's basically what the project aims were

T3: Cool

R: So yeah, what were your thoughts on the project aims to start with?

T3: I thought it was a really useful thing to find out because we do just tend to go into Year 7 and say "here's a test" and expect them to know what to do with it but I know, obviously, they have done SATS in year 6, but it's a very different style of questioning we don't teach them in any way how to answer those questions. So, it's, you know, something that I think is something long term something we should definitely be doing it's definitely, you know, I thought, a good idea to look into doing it with a small group of students and also to, you know, actually look at the outcomes. So we've done walking talking mocks with lots of other year groups and I know, obviously, it came from Pixel, I think, so they must have come up with some reason for doing it but we've never actually looked at the outcomes of whether they're different for students who had those style of exams. So yeah, interesting, but also potentially really important for those students' exam technique, you know, and now we're, kind of, looking to trying to drip feed all the KS4 stuff down into KS3, and we've always looked at content and never really looked at the exam technique; we do leave it quite late really to sort of teach them how to answer the questions. So yeah, I think, you know, really interesting but valid throughout KS3.

R: Yeah, and that was really the thoughts behind the project. So, the next question is how did you find the walking talking mock concept as a way of preparing students for exams?

T3: I have been doing it for a few years. I think I didn't do it the first year it was introduced because I'm not sure I was teaching the right year group at the time but I'm familiar for Year 11 and I've also used it in my own classes year 13 so yeah, I like it and I know how to do it, I've done it before. Possibly in not quite the same style as what we were doing with the Year 7s obviously it's a lot more prescriptive with the PowerPoint and all the rest of it. With year 13 I've just literally given them a paper and said 'this question is worth 6 marks, you've got 6 minutes and then we'll go over it at the end', so I think it was slightly different in the fact that we introduced it a lot more and the way it was structured pulled away the tablecloth gradually as you went through.

R: Yeah, and how did you find the walking talking mock went with your class?

T3: Pretty well actually, they're pretty receptive, they were a really good Year 7 class. Obviously, the usual range of ability in there but usually they were up for trying new stuff out, which was helpful. So, there wasn't any resistance to doing it, not that they, obviously, knew any better because they did it right from the start, but they were more than happy to take part in it. So yeah, it generally ran pretty well.

R: That's good! Did you think focusing on command words was the right approach to take during the walking talking mocks?

T3: Yeah, I mean, I don't know how else- I don't know what the other options were or how you could have approached it. I did find they got a bit fixated sometimes on those key words and potentially at the I don't know how sport wasn't what the other options are for how you could have approached it. I did find that they got a bit fixated sometimes on those keywords and potentially at the expense of reading the other parts of the question. Maybe, you know, the less able ones, and I don't know this is come into questions later, but I've got some thoughts about that. But I mean that was more about timings, but the keywords I think I think it's really important that they know them, I wonder if maybe there's something we should introduce before we sit an exam that there's something that we try to use more in lessons because they sort of, I don't know, became almost dependant on those keywords like highlighting like mad because I'd encourage them to highlight them and maybe they less able ones losing the kind of the concept of what they were meant to be using those command words for. I do in general think the process, I mean obviously you know that's what you're going to look at, you know, with the data as to how it affects different you know different types of students but I think the less able ones really liked it as a scaffold, particularly at the beginning and I think it probably suited them because the questions were done more slowly. I think it reduced a lot of anxiety because they were just given this great big long test and they didn't know where to start with it, and I think they probably were more willing to try questions that they maybe would have just ignored, done under normal circumstances, because they didn't understand them or they you know they thought it was too tricky. I think the more able students found the lack of flexibility a problem so they would have whizzed through the first couple of questions really quickly that would have given them more time to concentrate on the more difficult ones at the end of the paper and I think they didn't like that because, you know, it was so structured that they weren't able to manage their own time. So the students who could manage their own time would rather have done that, whereas the ones that couldn't were quite happy to be led.

****Full transcript available upon request****