



A multifaceted approach to improving higher order thinking skills in A-Level Biology

Megan Cooper

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

MSc in Learning and Teaching, 2024

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 allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.
5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:	Megan H D Cooper
Date:	23.9.2024

Abstract

This research aims to explore if an intervention can be implemented that impacts possible mechanisms for low attainment in questions requiring higher order thinking skills (HOTS). This small-scale practitioner study first aimed to establish the specific challenges experienced by a small group of Year 12 biology students when responding to such questions. Using multiple instruments, including assessments, questionnaires and interviews, a multi-faceted intervention was designed around strategies used to approach questions. The literature review suggested the effectiveness of explicitly modelling cognitive and metacognitive strategies, so a “think-aloud” intervention was implemented over a five-week period.

The impact of the intervention was evaluated by comparing pre-intervention to post-intervention assessments and through interviews with all participants. Findings suggest that the intervention had a positive impact on student self-efficacy, and cognitive and metacognitive strategies; however, this is not yet reflected in attainment. Students and teachers perceived the strategies introduced during the project to be useful but indicated a desire for less scaffolding and more feedback in future iterations. The study underlines the importance of metacognition and self-efficacy to science achievement and raises questions regarding how science attainment can be increased further.

1 CONTENTS

2	Introduction	5
3	Literature Review	8
3.1	Higher Order Thinking	8
3.1.1	HOTS and education – a drive towards change	8
3.1.2	Changes to curriculum and assessment	9
3.1.3	Historical perspective and definition	10
3.1.4	Development of HOTS.....	13
3.2	Suspected barriers to answering higher order questions	17
3.2.1	HOTS.....	17
3.2.2	Metacognition	18
3.2.3	Self-efficacy	18
3.2.4	Reading comprehension.....	19
3.2.5	Literacy - Scientific language	21
3.2.6	Content knowledge	22
3.3	Potential interventions	22
3.3.1	HOTS.....	22
3.3.2	Metacognition	24
3.3.3	Self-efficacy	26
3.3.4	Reading comprehension.....	27
3.3.5	Literacy – scientific language.....	27
3.3.6	Content Knowledge.....	28
3.4	Summary of Literature Review	29
4	Methodology.....	30
4.1	Overview	30
4.2	The participants	32
4.3	Collaboration	32
4.4	Instruments used to evaluate the research questions.....	33
4.4.1	Assessments	33
4.4.2	Student task during assessments.....	36
4.4.3	Student task after assessments.....	37
4.4.4	Questionnaire.....	38
4.4.5	Interviews.....	40

4.5	Ethical considerations.....	42
4.5.1	Consent.....	42
4.5.2	Power balance	43
4.5.3	Privacy	43
4.6	Limitations:	43
5	Findings and Discussion Part 1: Diagnosis.....	44
5.1	RQ1 - What are the barriers to answering higher order questions for my students?	44
5.1.1	Question types	44
5.1.2	Difficulty in question interpretation.....	48
5.1.3	Difficulty with the language of the question.....	50
5.1.4	Insufficient depth of knowledge, including a lack of biological vocabulary.....	52
6	Development of Intervention	53
7	Findings and discussion – part 2: Post-intervention	56
7.1	RQ2 – What impact does an intervention have in addressing the barrier to improve success in answering higher order questions?	56
7.1.1	Confidence.....	56
7.1.2	Structured approach shifts thinking.....	57
7.1.3	Attainment	59
7.2	RQ3 – What are student and teacher views about the process of diagnosis and intervention?.....	62
7.2.1	Test review using key	62
7.2.2	Perceptions of using the intervention strategies	65
7.2.3	Perception of teacher modelling.....	66
7.2.4	Conscious teacher planning	68
8	Conclusions and implications	69
8.1	Summary of Findings	69
8.1.1	Barriers for this cohort to answering higher order questions	69
8.1.2	Impact of the Intervention	70
8.1.3	Student and Teacher Views.....	71
8.2	Future Implications.....	72
9	References.....	74
10	Appendices.....	86
10.1	Appendix A : Full Version of Table 5.....	86
10.2	Appendix B: Questionnaire	87

10.3	Appendix C: Additional post-intervention questionnaire questions	88
10.4	Appendix D: Example of coding process used in analysis of questionnaire	89
10.5	Appendix E: Group interview questions	91
10.6	Appendix F: Example of coding used for interview questions.....	92
10.7	Appendix G: Questions from teacher interviews	93
10.8	Appendix H: Example of annotated question used to model think-aloud strategies during the intervention	94

2 INTRODUCTION

The ability to interpret and respond to examination questions accurately and appropriately is the key to success for candidates in written assessments. In education systems where external examinations are used to assess student performance, these assessments determine prospects and future pathways. Thus, there is a great deal of pressure on students to achieve to the best of their ability in these examinations and on teachers to help prepare students for the examinations.

As a teacher in the British education system, where many students are externally examined at 16 for GCSEs and again at 18 for A Levels, the underlying drive for daily teaching practice focuses on examination preparation in these classes. Lesson time is split between learning content, building skills and answering practice examination questions. Often, there is little time to deviate from the course specification and therefore a balance must be struck between teaching content and building skills. As a practitioner, each two-year course can feel time-poor, particularly in enabling students to develop the higher order thinking skills (HOTS) needed to answer the more challenging questions successfully.

Assessments have changed numerous times during my 17 years of teaching. In the most recent changes that took place between 2015 and 2019, the government emphasized the need for GCSEs

and A-Levels to become more “rigorous”, to move away from modular assessment, and to prepare students better for further study in their subject (Long, 2017; Office of Qualifications and Examinations Regulation & Department for Education [Ofqual & DoE], 2019).

From a practitioner perspective, students have found the changes at A-Level challenging, despite the claim from the UK government the demand of the course is the same (Ofqual & DoE, 2019). The difficulty for the students lies both in the change in content and a perceived increase in requirement to move beyond recall in the examination questions.

Following a set of relatively disappointing GCSE results in 2023, departmental analysis of the data found the cohort had underperformed particularly on higher order questions requiring more than recall of knowledge. This fits with my experience as a practitioner and highlighted a worrying trend. Success in examinations is important to the students, but also has implications for the future of my school, a co-educational independent day school. The question that arose as a consequence was: what could be done to address the lower attainment in higher order questions that stemmed from a lack of HOTS?

The literature suggests there has been a drive in education for centuries to improve HOTS in students (Resnick, 1987). While those in traditional, elite education had opportunities for this, the focus in publicly available education was on reading, writing and arithmetic, memorisation of facts from textbooks, and rote knowledge (Bransford et al., 2000; Liu et al., 2022; Resnick, 1987). The widely held belief was that complex understanding could arise only when foundational knowledge was achieved, thus there was no explicit teaching of critical reading, or complex problem solving in the early twentieth century (Liu et al., 2022).

The challenge of teaching HOTS to a larger proportion of the population emerged only after the advent of state education for all and the increasing demand of employers for more complex

skills (Bransford et al., 2000). In their systematic review of research on higher order thinking, Liu et al. (2022) found an upward trend in HOTS research internationally, particularly after 2014. This aligns with the call from policymakers and educators alike for increased teaching and assessment of HOTS. As a practitioner, however, pinpointing which aspect of HOTS to focus on in the classroom can be difficult, given that it is a complex term.

The aim of this practitioner research is to diagnose and address the underlying mechanism for low attainment when responding to higher order questions in my students. The project will be conducted with my small A-Level biology class. The size of the class affords an opportunity to collect rich data to better understand how to address the problem in context.

Questions to address in my literature review:

1. What are HOTS and their relation to education?
2. How are HOTS developed?
3. What does the literature say about the possible mechanisms for low attainment when responding to higher order questions?
4. What does the literature say about possible interventions to address these barriers?

3 LITERATURE REVIEW

The literature review begins with a brief overview of the importance of HOTS to education and how this relation has evolved over time, leading to an exploration of different ways of defining HOTS. This is followed by a look at the theories around how HOTS develop and how this relates to teaching. Based on long-term practice, several suspected mechanisms for low attainment in higher order questions are explored next, followed by an exploration of the evidence for potential interventions to address these barriers.

3.1 HIGHER ORDER THINKING

3.1.1 HOTS and education – a drive towards change

Many have decried the state of education and its lack of fitness for purpose, particularly in relation to a lack of HOTS. Before the turn of the century, Cuban (1984), pointed out that there were efforts throughout the twentieth century to develop students' thinking. US education at that time, "produce[d] a portrait of intellectually impoverished classrooms where unimaginative instruction tried to stuff a whale of knowledge into a sardine can of a student" (Cuban, 1984, p.655). Over the past few decades, there has been a movement from educators and policymakers alike towards reforming education to prepare students better for the rapidly changing world of the late-twentieth and early twenty-first centuries (Ben-Chaim et al., 2000; Cammies et al., 2024; Mkimbili, 2022; OECD, 2017; Osborne, 2013; UNEP, 2012; Zoller, 1993). Issues such as food security for a growing population, dealing with the consequences of climate change and ensuring sustainable energy sources are some of the most pressing problems of the twenty-first century identified by the United Nations Environmental Program (2012). To be able to face these challenges with scientific innovations, an updated approach by educational institutions is required. (UNEP, 2012).

While science education is valued internationally, the way in which science is taught does not necessarily reflect the changing needs of society. While Miri et al. (2007) point to a positive shift in science education from traditional, textbook-based rote learning to explorative, inquiry-based learning, rooted in real-world phenomena, Osborne (2013) disagrees that this produced the desired effect. He held the common rationale for science education in the twentieth century was economic in nature and focused on producing future career scientists. In doing so, the emphasis placed on foundational knowledge building at secondary school leads to conceptual coherence only at the undergraduate or graduate level. Those who do not pursue science further are left with a lack of skills, and accept learned content as irrefutable facts (Osborne, 2013). In our rapidly changing world, education is needed that enables all people to develop the intellectual tools, learning strategies, and knowledge necessary to think productively about different situations (Bransford et al., 2000).

3.1.2 Changes to curriculum and assessment

Many curricula and assessments around the world have been revised in response to the increasing demand for students to develop HOTS. For example, the PISA (Programme for International Student Assessment) 2015 science framework acknowledged some of the challenges faced by humanity in the twenty-first century by using the question, “What is important for young people to know, value and be able to do in situations involving science and technology?” to frame some of the changes (OECD, 2017, p. 20). In outlining the revised framework, it was recognised that most countries around the world value science education, but there are competing needs. Science education must prepare both those people who will progress to careers in science and educate those who will not take up these sorts of careers, but who nonetheless need to be informed citizens who are able to make choices and form opinions about science and technology (OECD, 2017). By focusing on developing scientific literacy in the revised framework, procedural

and epistemic knowledge are emphasized, in addition to content knowledge. With this, the PISA framework aims to meet the needs of developing both career scientists and informed citizens who can decipher scientific language and make judgements about information with which they are presented.

In England and Wales, GCSE and A Level courses were reformed between 2015 and 2019 along similar lines as the PISA assessment. The rationale, according to David Laws, the Schools Minister in April 2013, was to give students better preparation for post-16 study with the skills and knowledge needed to progress (Long, 2017). Presumably, this preparation was intended to prepare students for further study at university. As such, universities played a bigger role in the content reform at A-Level than previous iterations, with the aim of increasing rigour and allowing more time for learning, and less time spent in examinations (Long, 2017; Ofqual & DoE, 2019). Ostensibly, it can be argued that the additional time for teaching and learning would be used to develop the HOTS universities and employers value.

In the revised A-Level Biology course, the assessment objectives (AOs) are more concise compared to the previous iteration (AQA, 2021; AQA n.d.). All three AOs relate to scientific ideas, processes, techniques and procedures, but aim to assess different aspects of a student's knowledge. The first AO looks for a student to *demonstrate* their knowledge and understanding. AO2 aims to assess *application* of that knowledge and understanding, and the third AO is about *analysing, interpreting and evaluating* scientific information, ideas and evidence (AQA, 2021). Thus, students must be supported not only in their acquisition of content knowledge to succeed in AO1 and AO2, but also in HOTS to find success in AO2 and AO3.

3.1.3 Historical perspective and definition

Higher order thinking skills have been considered an important part of education at least since the time of the Greek philosophers, but different fields define the term differently. For Socrates,

Plato and Aristotle, critical thought and logic were held in high esteem and used to develop disciplined thinking (Lewis & Smith, 1993). More modern philosophers, such as Toulmin, teach general reasoning and argumentation skills to try to stem the “unhelpful” human nature of accepting faulty arguments and drawing unsubstantiated conclusions through undisciplined thinking (Resnick, 1987). Philosophers, therefore, emphasize reflective thinking, critical thought and logic as higher order thinking skills. In contrast, psychologists view higher order thinking as a cognitive process that is used in the construction of meaning and structure out of experience (Lewis & Smith, 1993). Higher order thinking produces behaviours that lead to problem solving (Cuban, 1984). Already the lack of clarity meant by HOT is apparent, given the inconsistency in definition between philosophers and psychologists.

HOTS, sometimes called higher order cognitive skills (HOCS), seems to be an encompassing umbrella term that is interpreted and used differently across disciplines and across the literature. There is a different focus for different academic disciplines. For example, critical analysis is the focus for English while mathematics and science focus on complex problem solving. In their systematic review of research on higher order thinking, Liu et al. (2022) found that other related aspects of HOTS included creative thinking (Hwang et al., 2018; King et al., 1998; Miri et al., 2007; Sun et al., 2022), metacognition (Hmelo & Ferrari, 1997; King et al., 1998; Sun et al., 2022), self-efficacy (Sun et al., 2022), and analysis, evaluation, synthesis and creation (Apino & Retnawati, 2017; Bloom, 1956; Anderson & Krathwohl, 2001), among others. Not surprisingly, many authors agree that defining HOTS is difficult and confusing (Cuban, 1984; Lewis & Smith, 1993; Resnick, 1987).

The term critical thinking is an important one to explore, as it is often used as an alternative or seen as an aspect of HOTS (Liu et al., 2022; Miri et al., 2007). According to Liu et al. (2022), critical thinking is the most frequently used phrase in research on HOT. Facione and Facione (2013) define

it explicitly as the combination of two things: the ability to analyse, evaluate and make inferences and the disposition to engage in critical thought processes. Yet even with Facione and Facione's (2013) decades of work on critical thinking, Lewis and Smith's (1993) observation arguably stands, that there remains confusion surrounding the term and its relationship to HOTS overall. While useful, critical thinking as defined and measured by Facione and Facione (2013) does not encompass the wider range of thinking skills employed when answering A-Level biology questions (Sun et al., 2022), so a broader definition is needed.

While recognizing the importance of critical thinking, Sun et al. (2022) propose HOTS are multi-dimensional and more constructivist with five constituents of HOT relating specifically to science. These include scientific reasoning, critical thinking, creative thinking, science self-efficacy, and metacognition. In a relatively large study, they looked at whether and how these constituents of scientific HOT relate to science learning and how the skills are interconnected (see Figure 1). Their results show that increasing metacognition and science self-efficacy, in particular, have strong effects on the other three aspects of scientific HOT (Sun et al., 2022).

The figure originally presented here cannot be made freely available via ORA because of copyright.

The figure was sourced at Sun, H., Xie, Y., & Lavonen, J. (2022). Exploring the structure of students' scientific higher order thinking in science education.

Thinking Skills and Creativity, 43, 100999.

<https://doi.org/10.1016/j.tsc.2022.100999>

Figure 1: Hypothetical model of relationships between five competencies that comprise science HOT (from Sun et al., 2022)

One thing upon which there does seem to be a consensus in the literature is that HOT goes beyond recall of facts and procedures. Many authors agree HOT is a form of thinking that is non-

algorithmic, complex, and which often generates multiple solutions (Miri et al., 2007; Resnick, 1987; Sun et al., 2022). There is an element of novelty, where the path ahead cannot be determined in advance, nor seen in its totality (Orton, 2004; Resnick, 1987). Nuanced judgement and interpretation must therefore be used to integrate, and possibly extend, prior knowledge and experience with new information to produce a solution to a previously unknown situation (Bartlett, 1958; Lewis & Smith, 1993; Maier, 1933; Orton, 2004; Resnick, 1987; Sun et al., 2022). HOT requires effort and depends on self-regulation (Resnick, 1987). For the purposes of this study, HOT will be defined as thinking that is complex and effortful, requiring integration of prior knowledge and new information to generate a solution to a novel situation.

3.1.4 Development of HOTS

While the idea that cognition increases in complexity from birth to adulthood is widely accepted, the way in which this happens is debated. In this section, the theories of Piaget, Vygotsky, and Bloom will be explored in their relation to the development of HOTS, followed by a discussion of how Bloom's theory is used in practice.

Piaget's cognitive constructivist view of learning, from which he developed the four cognitive stages of development, emphasizes the role that exploration and the interaction with the environment play in constructing meaning of the world (Palmer, 2005; Aubrey & Riley, 2022). The highest level of cognitive development, as exemplified by logical thought and abstract problem solving, develops in the final, formal operational stage (Piaget, 1977). The rigid, linear progression of Piaget's stages is limiting, however, as other researchers have found that some of the key thinking processes of one stage can appear in earlier stages (Aubrey & Riley, 2022). Equally lacking is the recognition that children develop in a social world and are influenced by those around them (Aubrey & Riley, 2022; Palmer, 2005).

Unlike Piaget, Vygotsky's constructivist view of learning, was formed around the social aspects of a child's experience (Vygotsky, 1978). His ideas around the influence of social and cultural factors on learning, scaffolding to promote movement from one "zone of proximal development" (ZPD) to the next and the importance of dialogue and language in the development of thinking skills have been widely influential in educational practice (Aubrey & Riley, 2022). Although criticised for being challenging to measure and for lacking in practical application, ZPD introduced the idea that a child's learning is influenced by a 'capable other' (Aubrey & Riley, 2022; Vygotsky, 1978). When that guidance is used appropriately, without an overreliance on help from adults or peers, a child becomes an active participant in progressing their learning.

Like Piaget's progressive cognitive stages of development, Bloom developed a taxonomy of cognitive learning objectives that is hierarchical and linear (Bloom, 1956). Students must achieve simpler objectives before progressing to the next, more complex one (see Figure 2). Each successive objective is increasingly difficult and is cumulative in nature, such that to comprehend an idea, knowledge must first be recalled. The six processes are often split into higher order and lower order thinking skills with knowledge, comprehension and application considered lower order thinking skills, and analysis, synthesis, and evaluation are HOTS (Irvine, 2017). The simple and straightforward nature of Bloom's method of categorizing thinking has been influential across many disciplines in planning and assessing learning. The levels therefore can be used to align intended learning outcomes with instruction and assessment (Aubrey & Riley, 2022; Irvine, 2017), aiding teachers in enabling their students to develop HOTS.



Figure 2: Bloom's taxonomy, showing progression from simpler to more complex cognitive educational objectives, constructed from Bloom (1956).

Categorising questions into Bloom's taxonomy levels can be used to help teachers and students identify which types of questions are found to be the most challenging (Cammies et al., 2024; Crowe et al., 2008; Zheng et al., 2008). In a study looking at higher education biology questions, Crowe et al. (2008) found that this enables teachers to align their instruction more closely to the needs of the students, thus improving student learning. Students are more likely to perceive teaching sessions that include HOTS, and which are constructively aligned with their assessments, as being more useful to their learning (Cammies et al., 2024).

Additionally, there is a positive impact on other aspects of thinking when instructors share the rubric used to classify questions, and when students receive specific feedback on performance. This is because students' metacognition is enhanced when they are better able to monitor their own learning (Crowe et al., 2008). Metacognition and its importance to the context of HOTS is recognized as an important aspect of learning and will be discussed further in Section 3.2.2.

Bloom's taxonomy has been criticized widely, however (Kreitzer & Madaus, 1994, Sun et al., 2022). It is seen as too simplistic to represent the complex nature of thought and learning (Furst, 1981). The cumulative, linear nature of the taxonomy is considered problematic as overlaps and inversions occur between the six objectives (Das et al., 2022; Furst, 1981; Jones et al., 2009; Kreitzer & Madaus, 1994; Marzano & Kendall, 2007). Additionally, there is disagreement as to which levels should be classified as higher or lower order thinking (Cammies et al., 2024; Crowe et al., 2008; Jones et al., 2009; Wineburg & Schneider, 2010; Zheng et al., 2008). For instance,

“application” is sometimes considered a transition between lower order thinking skills and higher order ones (Crowe et al., 2008). The lack of clarity and consensus around the hierarchy negatively affects the consistency with which it can be used.

To try to address the overly simplistic nature of Bloom’s taxonomy, Marzano & Kendall (2007) proposed a multidimensional taxonomy that accounted for metacognitive and “self” systems that work in conjunction with the cognitive domain. Marzano and Kendall’s taxonomy has had much less impact on the educational community, however, compared to Bloom or the revised taxonomy devised by Anderson and Krathwohl (2001; Irvine, 2017).

The revised taxonomy does not deviate greatly from Bloom’s original but provides more clarity and attempts to address the complex nature of thought. Anderson & Krathwohl provided clearer definitions of the six objectives of Bloom’s cognitive domain (Anderson & Krathwohl, 2001; Krathwohl, 2002). An additional knowledge dimension was introduced to separate knowledge as a noun from the six categories defined as verbs and renamed Bloom’s “knowledge” objective to “remembering” (Krathwohl, 2002). These changes went some way in constructing a more reliable system for planning and assessing learning outcomes, so individual educators could more accurately identify lower and higher items, an earlier problem summarized by Kreitzer and Madaus (1994). Yet, a more recent study by Das et al. (2022) highlights practitioners still find using the revised taxonomy a consistently difficult task.

One of the reasons using any of the outlined taxonomies in a consistent manner proves difficult is because of the flexibility of language (Stanny, 2016). A word can have different levels of meaning, influenced by the context in which the word is used. Therefore, in the revised taxonomy outlined above the verbs used to determine the taxonomy levels and thus to construct and assess learning outcomes are open to interpretation (Jones et al., 2009; Jordan, 1997). While several studies have sought to reduce the ambiguity in the cognitive level with which a verb is associated,

complete consensus has been deemed impossible, as some verbs have been found to occur in two neighboring levels, or multiple levels (Das et al., 2022; Stanny, 2016).

The problem with classifying objectives in a rigid hierarchy is arguably that higher order and lower order thinking skills are intertwined. Problems cannot be solved without the requisite content knowledge, skills and “relevant networks and structures in the mind” (Orton, 2004, p. 25). Knowing in an expert sense is more than recitation of a series of facts; it is an understanding of how concepts are interconnected. Even to perform what we think of as very “elementary” cognitive activities, such as reading, writing and solving some mathematical equations, requires complex thinking processes. For example, to understand a text being read, inferences must be made (Resnick, 1987). Conversely, existing domain knowledge has a large impact on acquiring new knowledge in an activity such as understanding and interpreting text (O’Reilly & McNamara, 2007). Thus, basic skills and higher order thinking skills cannot be separated easily during a process.

Perhaps, what emerges as useful from the collective work on taxonomies is a separation between words used to describe lower order thinking skills versus HOTS. As a practitioner, the classroom needs to be flexible and move fluidly between different skills; sometimes, categorising those skills can be useful. While there are many ways to define HOTS and categorise thinking, the goal is ultimately to increase the complexity of thinking.

3.2 SUSPECTED BARRIERS TO ANSWERING HIGHER ORDER QUESTIONS

3.2.1 HOTS

When HOTS are required to answer higher order questions, multiple skills will be used to generate a response (Sun et al., 2022). As outlined above, HOTS are complex. Any of these skills could be a barrier to answering higher order questions. While it is not practical to explore all these skills within the scope of this project, both metacognition and self-efficacy have been found to

strongly influence other HOTS in science (Magno, 2010; Sun et al., 2022). These will now be explored in turn.

3.2.2 Metacognition

While considered by some to be an aspect of HOTS (Sun et al., 2022), metacognition is widely recognised as a part of self-regulated learning (SRL). SRL is composed of cognition, metacognition and motivation. All aspects of SRL are thought to be important for learning, especially metacognition (Muijs & Reynolds, 2018). Metacognition is thought to be a person's "ability to think about their current cognitive processes" (Flavell, 1976, as cited in Sun et al., 2022). Schraw & Dennison (1994) break down metacognition into knowledge of cognition and regulation of cognition. There is a strong correlation between the two components, so what students know about themselves and available strategies can feed into how students plan, monitor and evaluate their learning (Schraw & Dennison, 1994). Furthermore, strong evidence exists relating metacognition to attainment (Muijs & Bokhove, 2020). If a learner lacks knowledge of their own cognition and is unable to regulate their cognition during a difficult activity, it is likely they will struggle with challenging activities, such as answering a higher order question.

3.2.3 Self-efficacy

Self-efficacy in science has been shown to have an influence on HOTS, such as critical thinking (Phan, 2009; Sun et al., 2022). Self-efficacy refers to a person's belief in their ability to perform a certain task (Bandura, 1977). Students who have a high self-efficacy in science will engage more frequently in tasks that require critical thinking, persevere through difficulties, and find more success (Linnenbrink & Pintrich, 2003). Conversely, if a student has low confidence in their ability to address more difficult questions in science, they are more likely to give up or avoid the question entirely (Bandura, 1993; Palmer, 2005).

Yerdelen-Damar and Pesman (2013) found a striking relationship between self-efficacy and metacognition. In their study of 338 10- to 14-year-old physics students, they reported that self-efficacy mediated the relation of metacognition to achievement in physics. Thus, without possessing the confidence to learn physics, a student's metacognitive skills did not independently mediate physics achievement (Yerdelen-Damar & Pesman, 2013). This finding gives credence to Sun et al.'s (2022) model of the interlinking nature of HOTS involved in science and Bandura's claim that perceived self-efficacy is central to academic development (Bandura, 1993; Sun et al., 2022).

Interestingly, high self-efficacy can be a barrier to learning as well. This can happen if a student has high confidence in prior knowledge leading to lower engagement in learning new ideas (Linnenbrink & Pintrich, 2003). Although this is not problematic if the student's knowledge matches their belief, it can limit necessary conceptual changes (Linnenbrink & Pintrich, 2003).

3.2.4 Reading comprehension

Answering higher order questions first requires the student to decipher the text of the question. This involves recognising the words used, interpreting the context and inferring meaning. Readers activate schema to access prior knowledge and experiences, which are important to build meaning from the text (Pardo, 2004; Vacca et al., 2014). Schema are the way in which prior knowledge and experiences are organised and stored in the long-term memory (Anderson & Pearson, 1988; Pardo, 2004; Vacca et al., 2014). When schema are activated, they help build meaning from text in a number of ways. For example, schema help select relevant information from the text to make inferences. They integrate the new information in the text with prior knowledge. Finally, they elaborate information so that more insightful judgements and evaluations can take place (Vacca et al., 2014).

The way in which a reader activates their schema depends on the level of comprehension used, and thus level of challenge. The empirical evidence provided by Alonzo et al. in their 2009

study supports the theory that there are three levels of comprehension when a reader interacts with the text: literal, inferential and evaluative. Their study, looking at 300 third graders, found the literal level of comprehension is easier for students than both the inferential and evaluative. While this finding is not surprising, it gives weight to the idea that making inferences and evaluating require a higher level of cognitive skills. The authors recognise the limitations of their findings, noting comprehension is a complex task which probably requires a non-hierarchical interaction of all three levels of comprehension (Alonzo et al., 2009). The way in which the three levels interact depends on the level of abstraction inherent to a question.

When faced with text from a different content area, such as science, students might not possess the understanding, prior knowledge, or thinking skills to enable evaluative comprehension in that domain (Jackson, 2016). The importance of reading skills in a content area is highlighted in a study by O'Reilly & McNamara (2007). They found that reading skills impact science achievement for students who possess either low or high domain knowledge. High levels of reading skills can even compensate for low domain knowledge (O'Reilly & McNamara, 2007). This does not diminish the importance of domain knowledge, nor the interaction between these factors. Those students with both high reading skill and high knowledge had much higher course grades, compared to others that had only high reading skill or high knowledge.

Metacognitive reading strategies also contributed significantly to the final course grade but proved a weak predictor of science achievement. The difference between these two outcomes is possibly related to the more complex, long-term nature of a course which may contain more complex and "authentic" assessments of science knowledge than the standardized assessment used to measure science achievement in the study (O'Reilly & McNamara, 2007). The conclusion drawn by the authors is therefore arguably weak, given the variance in teacher practice and the different factors which contribute to course grades.

3.2.5 Literacy - Scientific language

In their guidance document on “Improving secondary science”, the EEF state, “learning science involves learning a whole new language” (Holman & Yeomans, 2018, p. 30). The domain-specific vocabulary of science is often unfamiliar, formed from multiple morphemes, and refers to abstract concepts (Fisher & Blachowicz, 2013; Holman & Yeomans, 2018). Students presented with a scientific text are faced with language designed to convey complicated ideas efficiently, tending to be very formal, impersonal and complex (Snow, 2010). Given the lack of similarity this bears to the informality of casual conversation, it is not surprising adolescents struggle to decipher these terms (Fang, 2006; Snow, 2010). In addition to the barrier formed by the need to recall or decipher unfamiliar scientific language is the expectation to use the vocabulary in oral and written discussion in a way that justifies the EEF’s statement that learning science is like learning a language.

Not only do students need to learn and understand the domain-specific language, they are also faced with non-domain specific terms that frequently appear in the text but are difficult to define using context alone. These “Tier II words”, such as “compress”, “random” or “abundant”, are often not recognized or understood (Beck et al., 2013; Pickersgill & Lock, 1991; Vacca et al., 2014). If a student encounters an unfamiliar Tier II word, they might struggle to decipher the meaning of the sentence or paragraph, even in the context of a familiar concept.

Even if a student manages to decipher unfamiliar words, the use of more complex language can increase cognitive processing time (Brown et al., 2019). In their study looking at the effect of complex scientific language on working memory capacity, Brown et al. (2019) found students were consistently slower in making decisions when complex language was used during instruction. The reduced memory capacity can increase “forgetting” and become a barrier to conceptual learning (Brown et al., 2019). In an examination, the use of complex language in

questions creates an additional barrier to accessing a question and increases the time pressures to respond.

3.2.6 Content knowledge

It is generally recognised that knowledge is essential to more complex understanding and success in school science (Holman & Yeomans, 2018). Domain-specific knowledge is required to be able to analyse, interpret, evaluate and use new information to solve a novel problem (Newmann, 1990; Zoller, 1993). Many believe that when acquisition of basic factual knowledge is secure, higher order learning can then take place (Bloom, 1956; Willingham, 2021). Having secure factual knowledge reduces cognitive load and the demand on working memory, freeing more cognitive load for problem solving (Sweller, 2010). Science knowledge has an impact on both science achievement and how well new information is learned, yet students often fail to draw on their knowledge appropriately or completely when solving problems or learning new information (O'Reilly & McNamara, 2007).

3.3 POTENTIAL INTERVENTIONS

3.3.1 HOTS

With the promise of HOTS as a key indicator of success (Lee & Choi, 2017) studies have been carried out to see whether HOTS can be taught directly, with varying success. A longitudinal study by Miri et al. (2007) identified three teaching strategies linked to potential improvement in critical thinking skills and disposition over three years. These strategies included using interdisciplinary real-world scenarios, open-ended class discussions and short inquiry experiments carried out in groups, all with the express purpose of fostering critical thinking in their students (Miri et al., 2007). Only two out of ten teachers in Miri et al.'s (2007) study used these strategies, however: a finding consistent with other evidence that many teachers use a more traditional, concept-driven approach to their practice (Mkimbili, 2022; Newmann, 1990; Osborne, 2013).

Even if teachers design their instruction to foster HOTS, there is a prevailing belief that only higher-achieving students will benefit. A series of studies reported on by Zohar and Dori (2003) found both high and low-achieving students benefit from instruction designed to foster HOTS. This finding, in conjunction with the implication from Miri et al.'s study that many teachers' practices are concept-driven, suggests teachers may require professional development before endeavouring to teach HOTS in their classroom.

One approach to improving HOTS is by direct instruction of problem-solving strategies, using a heuristic approach. This approach teaches problem-solving skills which can then be applied to other problems (Muijs & Reynolds, 2018). Pólya (1957) outlines a four-step sequence in the process including understanding the problem, devising a plan, carrying out the plan and evaluating the plan. Establishing both linguistic understanding and an understanding of the problem is a necessary first, rather than making too quick a judgement about what the question is asking (Muijs & Reynolds, 2018). In a meta-analysis of studies on problem-solving in mathematics education, Hembree (1992) found that problem representation, such as drawing a diagram, had large effects on performance. Asking pupils to explain each step increases problem-solving success (Gagné 1965, as cited in Muijs & Reynolds, 2018). Different heuristics can be explicitly taught for different situations.

Much work has been done on the heuristic approach in mathematics education, particularly on solving word problems (Muijs & Reynolds, 2018), but the approach can be applied to other domains (Schraw et al., 2006). A heuristic approach has been shown to be effective, particularly in older secondary school students and when the teachers providing the direct instruction had been trained in heuristic methods (Hembree, 1992). Hembree postulated the link between the impact of heuristic instruction and age, which is possibly linked to cognitive development and the age at which students develop abstract thinking (Hembree, 1992; Piaget,

1977). In contrast to this, Schraw et al. (2006) found that strategy instruction is most helpful to younger students and those under-achieving, particularly when it combines several interrelated strategies. These contrasting empirical studies point to the possibility that all ages can benefit from problem solving instruction, if the appropriate strategies are chosen.

In teaching strategies directly, it would be the hope of many teachers that students would be able to transfer these strategies from one problem to the next, from subject to subject and from one academic year to the next (Bransford, et al., 2000). With each novel problem solved, a person would presumably have learned something new, which can then be applied to other, similar problems (Orton, 2004). In contrast, Gagné (1965, as cited in Orton, 2004) argues that problem-solving cannot be taught. While individuals can learn something from solving a problem, they do not necessarily become better at solving problems. Each novel problem will require slightly different skills (Orton, 2004). Muijs and Reynolds (2018) agree the problem with both heuristic and metacognitive methods is transfer. They argue students will revert to randomly applying algorithms or guessing when faced with new situations, but suggest strategies can be used to better enable transfer (Muijs & Reynolds, 2018).

3.3.2 Metacognition

Considerable evidence has accumulated to show improved metacognition supports learning in science (Holman & Yeomans, 2018; Sun et al., 2022). The EEF recommends improving self-regulation strategies, particularly metacognition, as one of seven recommendations to benefit

The figure originally presented here cannot be made freely available via ORA because of copyright.

The figure was sourced at Quigley, A., Mujis, D., & Stringer, E. (2021). Metacognition and self-regulated learning: Guidance report. Education Endowment Foundation. EEF_Metacognition_and_self-regulated_learning.pdf (d2tic4wvo1iusb.cloudfront.net)

Figure 3: Relationship between cognition and metacognition from Quigley et al., 2021

secondary science (Holman & Yeomans, 2018). As the cyclical relationship between metacognition and cognition involves several stages, there are many potential foci for strategy instruction. The complexity of the model (simplified in Figure 3) means it would be challenging to focus on all parts at once. Whichever focus is chosen, the EEF recommends the instruction of metacognitive strategies should be explicit, delivered in a subject-specific manner and as part of a specific task (Holman & Yeomans, 2018; Quigley et al., 2021).

Modelling by the teacher can be useful for explicitly demonstrating a skill or behaviour, which can be linked to prior skills and knowledge (Muijs & Reynolds, 2018). Modelling thinking processes provides a multi-faceted approach to model cognitive and metacognitive strategies to demonstrate strategy selection, rationale for selection and evaluation (Holman & Yeomans, 2018). These think-aloud processes allow “expert” thinking, which is usually hidden, to be observed by the students (Rosenshine & Meister, 1992).

Modelling thinking can be particularly useful when approaching a new problem or way of thinking, followed by scaffolded student practice of the modelled strategies (Holman & Yeomans, 2018; Muijs & Reynolds, 2018; Rosenshine & Meister, 1992; Vacca et al., 2014). Through scaffolding, students start to internalize a process and can work to apply their new skills independently (Jackson, 2016). As students practise thinking aloud, they are actively participating and assessing their own strategy use (Rosenshine & Meister, 1992). Teachers can then appraise which cognitive processes their students are using (Jackson, 2016) and provide feedback. A study conducted with young primary school children by Bohn et al. (2004) emphasized the importance of providing positive feedback to those who were modelling well. This will inform those students that they are finding success in building their metacognitive and cognitive skills.

Scaffolding is seen as an important part of the constructivist method of teaching, whereby all knowledge is constructed and built from prior knowledge and experience (Muijs & Reynolds,

2018; Vygotsky, 1978). The scaffold provides the framework for the student to progress to the next level of mastery (Dabarera et al., 2014) and helps “to bridge the gap between current abilities and intended goal” (Rosenshine & Meister, 1992, p. 26).

The difficulty of a scaffolded task must be carefully regulated to ensure that students have sufficient background knowledge and skills to be able to progress (Rosenshine & Meister, 1992; Vygotsky, 1978). Bjork and Bjork (2014) define desirable difficulties as those that “trigger encoding and retrieval processes that support learning, comprehension and remembering” (p. 62). A balance is found whereby students are both challenged and possess the wherewithal to respond to these challenges successfully.

3.3.3 Self-efficacy

Bandura proposes four factors to improve perceived self-efficacy: mastery experiences, vicarious experiences, verbal persuasion and physiological/affective states (Bandura, 1977; Palmer, 2005). Mastery experiences are those in which a student experiences success when dealing with a challenge. To provide challenge and enable success for more mastery experiences, teachers need to ensure tasks have desirable difficulty (Bandura, 1997; Bjork & Bjork, 2014; Palmer, 2005). Student self-efficacy also can increase by watching a behaviour modelled by another person, which can be a teacher or a peer (Palmer, 2005). Receipt of positive feedback is the third source of self-efficacy. This verbal persuasion should be focused on effort as well as achievement and ensure it engenders self-efficacy that is an accurate reflection of ability (Linnenbrink & Pintrich, 2003; Palmer, 2005). Lastly, a student’s physiological/affective state refers to a student’s interpretation of their own stress levels. Interpreting anxiety as a normal response, rather than as a sign of inability, leads to higher self-efficacy (Palmer, 2005). Therefore, teaching practice which targets any of these factors may help increase self-efficacy.

3.3.4 Reading comprehension

Direct instruction can provide a useful intervention when developing reading comprehension (Jackson, 2016). Jackson looked at the impact of a think-aloud intervention on the evaluative comprehension skills of 48 first graders in relation to science-content texts. Those in the intervention group showed a significant improvement in evaluative comprehension when compared with the control group (Jackson, 2016). While the findings here show a promising intervention for reading comprehension in a domain-specific context, the small size of the study and the age of the students limits the generalisability of the findings, particularly to older students. O'Reilly & McNamara's (2007) work with older students points towards the possibility of an intervention that teaches reading strategies and skills explicitly, although further research is needed on whether these reading strategies impact science achievement directly.

Strategies to improve one aspect of learning often link to another. Allen and Hancock (2008) found that metacognition positively affects reading comprehension. In their study on 196 4-6 graders, Allen and Hancock trialed the use of individualized cognitive profiles of skills related to reading comprehension as measured by the Woodcock-Johnson III test. The profiles were shared with the students with relative strengths and weaknesses presented in a graphical format. During a comprehension activity, the students could actively monitor and mediate their skills by using their own strengths to overcome their weaknesses; thereby, constructing meaning from the text and context (Allen & Hancock, 2008). The findings here suggest that the importance of metacognition in its relation to the development of HOTS cannot be underestimated.

3.3.5 Literacy – scientific language

A significant number of studies have found literacy strategies, both narrow and broad, improve academic achievement (Beck et al., 2013; Brown & Concannon, 2016; Nunes et al., 2017). Nunes et al. (2012) reported the use of morphemes in decoding was a strong predictor of reading competence in students. Using this work, the EEF recommends explicitly teaching vocabulary using

morphemes so new words will be recognised more easily in the future (Holman & Yeomans, 2018). Rather than focusing on the use of one narrow strategy, Brown and Concannon (2016) found using multiple close reading and vocabulary strategies positively affects students' perceptions of vocabulary knowledge and their content achievement. The authors suggest that the positive effect the strategies had on students' perceptions was linked to their metacognitive ability to select appropriate strategies when a lack of understanding was recognised during reading (Brown & Concannon, 2016). This implies direct instruction of literacy strategies, supported by increased metacognitive awareness, can have a substantial impact on understanding the language used in science and scientific assessments.

3.3.6 Content Knowledge

To improve knowledge and understanding of science content, the EEF recommends embedding practice that enhances long-term memory formation and retrieval (Holman & Yeomans, 2018). Teachers are urged to use strategies that take cognitive load into account and find ways to reduce extraneous load, enabling students to focus on materials relevant to the task (Sweller, 2010; Torrence Jenkins, 2017). Karpicke & Roediger (2008) found retrieval practice hugely beneficial to long-term retention of knowledge. The more knowledge is retrieved, the stronger the schema become in our brains, thus reducing the cognitive load and the demand on our working memory (Agarwal, 2019; Holman & Yeomans, 2018). Spacing retrieval practice across time enhances this further (Holman & Yeomans, 2018). However, Agarwal (2019) cautions that learning cannot be solely focused on fact-based retrieval to achieve higher order learning. He found higher order retrieval practice, or a mixed retrieval practice, increased higher order learning (Agarwal, 2019). This enabled both younger secondary students and tertiary students to address higher order questions more successfully.

3.4 SUMMARY OF LITERATURE REVIEW

The review of the literature confirms that knowledge acquisition and the development of HOTS is highly complex. While educators and policymakers alike agree that it is important to develop HOTS, there is no clear consensus on how HOTS develop. Since HOTS encompass such a broad range of skills, there are numerous methods for engendering them in students. From a practitioner perspective, context is also important, and each student will face individual challenges and successes as they develop HOTS. Thus, it is important to me that my investigation focuses on the needs of the participants, developing an intervention around the barriers with which they contend specifically.

Deciphering higher order questions adds an additional dimension to employing HOTS. When answering higher order questions, students encounter several potential barriers; however, the literature points to both metacognition and self-efficacy as important factors to consider (e.g. Sun et al., 2022). In my investigation, it feels prudent to include both an aspect of metacognitive reflection and assess the degree to which perceived self-efficacy is impacted.

Therefore, the research questions for the project are:

- 1) What are the barriers to answering higher order questions for my students?
- 2) What impact does an intervention have in addressing the barrier to improve success in answering higher order questions?
- 3) What are student and teacher views about the process of diagnosis and intervention?

4 METHODOLOGY

4.1 OVERVIEW

This small-scale practitioner research project was carried out with my Lower Sixth teaching class (Year 12 – age 16-17). Using case study approaches in the context of my practitioner research enabled the collection of multiple, overlapping sources of data to increase credibility (Tracy, 2010). The rationale being that a small study, looking at a phenomenon in its real-world context, would provide the potential for a more detailed understanding that would not be possible in a broader study (Denscombe, 2021; Yin, 2018).

To answer the research questions, a stepwise approach with two phases was undertaken. The first phase was designed to diagnose the challenges experienced by students when responding to higher order questions. From the challenges identified, a specific focus was then chosen for a support intervention, the development of which is discussed after the findings from RQ1 are presented. Implementation of the intervention and evaluation of its impact formed the second phase.

A mixed-methods approach was developed to enable the collection of rich data that could be triangulated to provide a holistic view of the problem (Cohen et al., 2018; Denscombe, 2021). Quantitative data was collected from assessments to measure attainment for different question types and from closed questions on a questionnaire. Qualitative data was collected from questionnaires and interviews to complement the assessment data, and the richness of the data allowed for a deep and nuanced understanding of the context (Menter et al., 2011). Quantitative and qualitative data were given equal emphasis. A summary of the research project can be found in Table 1, giving a brief description of the research activities undertaken and the purpose of each one.

Duration/ weeks	Phase of Research	Research activity	Purpose
7	Diagnostic phase	End of topic assessments	<ul style="list-style-type: none"> To measure overall attainment To measure attainment related to different question types To identify the types of questions students perceive as most challenging
		Questionnaires	<ul style="list-style-type: none"> To measure student confidence To acquire student perceptions about question challenge and barriers to success
		1:1 Interviews	<ul style="list-style-type: none"> To explore students' reading comprehension To explore thinking processes To gain teacher perspective on barriers to answering higher order questions
5	Intervention	Intervention Introduction	<ul style="list-style-type: none"> To introduce cognitive strategies and the ideas around metacognition Model the think-aloud process
		Summative assessment and questionnaire	<ul style="list-style-type: none"> School-wide assessment point Compared data to pre- and post-intervention data.
		Intervention	<ul style="list-style-type: none"> To practise using cognitive and metacognitive strategies using a scaffolded think-aloud process.
1	Post-intervention	Post-intervention end of topic assessment	<ul style="list-style-type: none"> To compare attainment in question type to pre-intervention attainment To determine changes in student perception compared to pre-intervention
		Post-intervention questionnaire	<ul style="list-style-type: none"> To explore student perception of and impact of intervention
		Group Interview	<ul style="list-style-type: none"> To further explore student perception and impact of intervention, with flexibility of conversation
		1:1 teacher interview	<ul style="list-style-type: none"> To ascertain any changes to teacher perception from pre-intervention interview. To explore impact of intervention on teacher practice.

Table 1: A summary of the phases of the research project

4.2 THE PARTICIPANTS

The project was carried out with my current Lower Sixth teaching set. The class contains five mixed-ability students: four females, one male. A non-examination year group was chosen purposely so the progress of the project would not be impeded by the timings or pressures of external examinations, nor would examination preparation be impeded by the research project. They are the only set of Lower Sixth A-Level biologists at the school, so no comparison group was possible in the context.

The teacher participant was also a collaborator, chosen because he was the co-teacher of the participating Lower Sixth class. This enabled more data to be collected from assessments and provided additional opportunities to enact the intervention. His unique perspective as another teacher of the class and the wealth of explicit and implicit practitioner knowledge he possesses provided a valuable tool to measure the impact of the intervention on the students and on his practice (Jarvis, 1999).

4.3 COLLABORATION

I collaborated closely with the other teacher of the class, meeting several times to discuss the development of the intervention. In these meetings I shared my initial data, ideas, and some theoretical backing after which we discussed how we could refine the intervention. Our dialogue allowed us to combine our knowledge of the participants, the curriculum, and teaching experience, thus influencing my ideas for the intervention (Cordingley, 2015; Jarvis, 1999). The shared risk-taking involved in using a new approach enabled us to support one another through the process of implementation (Cordingley, 2015).

I also collaborated with a colleague in the Learning Support department in my school when designing the interviews to assess reading comprehension. She has responsibility for carrying out various assessments, including assessing reading comprehension. She confirmed that my questions

were appropriate to measure reading comprehension. She felt they asked the students to retrieve detail and infer meaning of words in the context of the text.

4.4 INSTRUMENTS USED TO EVALUATE THE RESEARCH QUESTIONS

4.4.1 Assessments

This instrument was used to evaluate RQ1 and RQ2. The research phase, topic, duration and space in which the six assessments used for the project took place are outlined in Table 2. The assessments during the intervention phase were part of a school-wide assessment procedure and were not explicitly part of the project. The data are included, as the assessment fell within the project timeframe.

Research phase	Assessment topic	Duration (min)	Took Place in:
Diagnostic	Cell membranes	40	My lessons
	Immunity	40	
	Ecology	40	Colleague lesson
Intervention	Mixed	90	Examination Hall
	Mixed	90	
Post-intervention	Variation and selection	40	My lesson

Table 2: An overview of information about assessments

For each assessment, past examination questions were selected to cover a range of question types and concepts within the assessed topics. The intention was to make each assessment close to what would normally happen in lessons and still provide a range of question difficulty for the purposes of the project. Particular attention was paid to including questions that required interpretation of data or diagrams as, from my experience as a practitioner, these often prove among the most difficult for students.

Both overall attainment and attainment per question were measured and analysed. Calculating overall attainment as a percentage allowed comparison between assessments that had a different number of marks. Individual attainment was analysed over the sequence of the

assessments outlined in Table 2 to look for any change in attainment, particularly following the intervention to evaluate RQ2.

As the number of marks available for each question varied, each response was ranked according to the marks attained as seen in Table 3. The ranking system enabled attainment for each question to be standardized and compared across questions and across tests during analysis.

Question Attainment Rank	Marks for response	Description
1	Full marks attained	Demonstrated sound knowledge, understanding and use of key terminology. Met the demands of the question.
2	Partial marks attained	Minor mistakes indicating partial understanding, a lack of key terminology or missing detail.
3	Response given, but no marks attained	Response not scientifically accurate. Major mistakes. Did not meet demand of question.
4	No attempt	

Table 3: An explanation of the ranking system used to rank attainment for each assessment question

There is a limitation in using the above ranking system for question attainment, as nuance was lost from the middle rankings. The reasons for gaining only partial or no marks can vary considerably. More may have been gained from a deeper, more detailed analysis of each student response. Time was limited during the diagnostic phase, however, as the data needed to be used to decide the direction of the intervention. Thus, I decided to use a less nuanced system, acknowledging the limitations of conclusions I can draw from the data.

To identify which questions both the cohort and individuals found most challenging for each assessment, mean attainment rank was calculated per question and per student. For analysis, questions were then split into three categories based on the mean attainment rank, as outlined in Table 4. A “Question difficulty” level was assigned to each category to relate student outcome to how difficult the cohort found a question. Using this system, a less difficult question indicates all

students achieved at least some marks. For a more difficult question, at least two out of the five students did not gain any marks and no more than two students would have gained full marks.

Mean attainment rank	<i>Question difficulty</i>
<2.1	Less difficult
2.1 – 2.5	Moderately difficult
>2.5	More difficult

Table 4: Question difficulty based on mean question attainment rank

As part of my analysis of the data for RQ1, each assessment question was further organised by question type within each question difficulty category. Six general question types were identified from the command words used, exemplified in Table 5 (full table in Appendix A). Parallels can be drawn between these six question types and some of the learning outcome categories used in the revised taxonomy (Anderson & Krathwohl, 2001; Krathwohl, 2002). Despite the recognised ambiguity of the revised taxonomy (Stanny 2016), I cross-referenced my groupings with the action verb classification done by Das et al. (2022) as a framework for comparison. If there was more than one command word in a question, it was placed in the category which I felt had a higher cognitive demand. In the analysis, frequencies of each question type were counted, and the distribution of different question types was compared across the three question difficulty categories to see if there was a pattern.

Question type	Demand of question (AQA, 2014)	Link to revised taxonomy using action verbs (Das et al., 2022)	Examples of command words included in this category
Making a judgement – using evidence	Making an informed judgement; judging from available evidence or supporting a case with evidence.	Evaluate	Is the conclusion correct? Justify Assess why
Suggest	Present a possible cause		Suggest and explain Use the data to suggest how
Explain	Give reasons	Comprehension	Explain how Describe and explain
Dealing with data		Application	Calculate Describe relationship (on graph)
Describe	Give an account of	Knowledge & Comprehension	Describe what Give and describe
Recall		Knowledge	What State (use table)

Table 5: Examples of categorisation of examination question command words into question type

4.4.2 Student task during assessments

This instrument was used to evaluate RQ1.

During the first three assessments, students were asked to indicate the three questions they found most challenging and provide an explanation for their choice. Their notes were used in conjunction with and compared to data from a post-assessment reflection afterwards. The intention was to explore student perceptions of challenge close to the moment of the assessment. It is recognised that the metacognitive nature of the student's reflection altered thought sequences that would have occurred during problem-solving (Ericsson & Simon, 1993). The data were used to provide an initial picture as to what barriers the students face when answering questions during a test situation.

During analysis, frequencies of the types of question selected by the students were counted and were compared to question attainment. This allowed me to see if there was a pattern in the type of question selected and to see if perceptions of difficulty aligned with attainment. The data set is incomplete as not all students elected to complete this task, or difficult questions were identified without providing an explanation of choice. As it was important to me that their response was recorded during the assessment, I did not follow up the null responses afterwards.

4.4.3 Student task after assessments

This instrument was used to evaluate RQ1.

Following each assessment, a 50-minute lesson was used to review the test. During the first of these lessons, the students and I collectively devised a coding system for reasons they did not gain full marks in questions, as outlined in Table 6. Students used this system as they reviewed each test to code the questions for which they did not gain full marks. The code they chose was based on their perception of the barrier(s) they faced and links to the evaluative stage of the metacognitive cycle (Muijs & Bokhove, 2020). This coding key is similar to those I have been using as a practitioner for years.

Code	Description
V	Didn't use specific vocabulary
K	Lack of knowledge
U	Lack of understanding of topic
Q	Lack of understanding of question
D	Not enough detail provided in response
M	Mathematical skills lacking
S	Silly mistake

Table 6: Coding key devised to indicate why full marks were not attained for a question

During analysis, frequencies for each code were counted and the distribution of codes was compared across the question difficulty categories. As part of my evaluation of RQ1, this allowed for an exploration of a possible relationship between question difficulty and student perceptions of why they did not achieve more marks. If a relationship was found to exist, it might highlight a salient barrier, around which to focus my intervention. Secondly, frequencies for each code were looked at in relation to each student. I was interested to see how the individual challenges experienced by different students were expressed in their choices. Not all students elected to assign a code to questions, however, and some students assigned up to three codes to a question.

4.4.4 Questionnaire

This instrument was used to evaluate RQ1, RQ2, and RQ3.

After the assessments, questionnaires were distributed online via Microsoft Forms (the platform used by my school) to elicit students' perspectives on confidence and on their performance in relation to answering higher order questions (see Appendix B).

A five-point Likert scale was chosen to measure confidence in relation to five statements based on a similar questionnaire used by Walsh et al. (2021) to measure nurses' confidence in challenging situations. While nursing and education have inherently different challenges, Walsh and her colleagues suggest their scale could be modified and used as a tool for other disciplines (Walsh et al., 2021). Descriptors were used rather than both numbers and descriptors, so students did not need to process two pieces of information to make a decision (Cohen et al., 2018). The various levels on a sliding scale can be open to interpretation; however, five points have been shown to provide reasonable reliability, without overwhelming the participant with choice (Cohen et al., 2018). Rankings of the confidence descriptors were assigned afterwards to enable easier analysis, as outlined in Table 7.

Confidence descriptor	Assigned rank
High Confidence	5
Confidence	4
Some confidence	3
Little confidence	2
No confidence	1

Table 7: Confidence descriptors and assigned rank in Likert scale

The confidence rankings were used to evaluate RQ2 to see if students' confidence changed in relation to various factors over the course of the study. Data were analysed by looking at the trend for each student, and for the cohort overall, for each statement. Increased confidence would indicate a student's self-efficacy had increased in that measure, and self-efficacy can be linked to HOTS in science (Sun et al., 2022; see discussion above).

To further evaluate RQ1, four open questions were included. Together with the data collected in the post-assessment coding, the two sets of data provided a more complete picture of students' perspectives and the context around their performance (Denscombe, 2021).

The open questions were similar to those often used in my lessons as part of a post-assessment reflection, so the students were familiar with the process (see Appendix B). The first question allowed for an exploration of students' perceptions about the converse of what they found challenging. Questions 2 and 3 allowed for elaboration of data from the students' post-assessment coding. I was also interested in exploring whether my intervention could be informed by students' perceptions of how they might improve. Thus, the fourth question intended to enable the evaluative stage of the metacognitive cycle as students reflected on their performance and looked towards improvement (Muijs & Bokhove, 2020). In the post-intervention questionnaire,

two additional questions were added (see Appendix C) to allow evaluation of RQ2 and RQ3. These questions elucidated students' perceptions of how and why their confidence changed since the start of the study.

Content of the responses was analysed in different ways to find meaning. The responses were coded to look for the mention of key concepts. Similar concepts were grouped together, and patterns were sought across each question per student and per assessment (Cohen et al., 2018). See Appendix D for an example of coding across a question.

4.4.5 Interviews

All interviews were recorded and transcribed verbatim, then changed for readability without losing the original meaning.

4.4.5.1 *For Reading Comprehension*

This instrument was used to evaluate RQ1.

As reading comprehension was one of the suspected mechanisms for low attainment, I wanted to explore this in a focused manner. Past examination questions were selected that included longer passages of text. As I was interested in the evaluative comprehension of reading, I developed interview questions that related to understanding the selected passages rather than using tests such as NARA II and TOWRE that measure different aspects of reading comprehension (Nunes et al., 2012). The meaning of the command words was probed to gauge whether the students understood the action verb.

During analysis, student responses were judged on accuracy in relation to retrieval of information from the text and inferred meaning based on my domain knowledge and the mark schemes to the questions. The transcript and my field notes were used to infer time taken to respond and other non-verbal aspects of their response.

4.4.5.2 *To Explore HOTS*

I planned and began carrying out think-aloud interviews for the purpose of exploring students' thinking. Practice questions and procedures recommended by Leighton (2017) were followed (see Appendix E). When one of the students became visibly upset during the process and felt she could not continue, I decided to stop. From the outset, the student struggled to verbalise her thoughts. Gentle prompts to 'keep talking' unsettled her. She expressed that she did not know the chosen topics well enough to answer and started to apologise. The power asymmetry in our relationship in the context of a one-on-one interview may have created a tension between wanting to stop the interview and a fear of disappointing me. Even if a think-aloud interview is presented as low-stakes, performance anxiety can increase and affect performance (Leighton, 2017). It felt unethical to conduct any further think-aloud interviews, as I did not want to risk creating further undue stress. I have not included the data as part of my research project

4.4.5.3 *Group interview*

Following the intervention, a semi-structured group interview was conducted with the cohort during a 50-minute lesson. The interview method was chosen to allow for flexibility of conversation and opportunities for clarification from all parties. Dialogue between members of the group was possible, so ideas could be built upon in a way not possible in individual interviews (Cohen et al., 2018; Denscombe, 2021). Prior to the interview, I printed the indicative questions for the interview and provided ten minutes at the beginning of the lesson for the students to make notes on the questions. This counteracted the potential pitfall of a member of the group potentially dominating the conversation and it reduced potential anxiety about responding to questions spontaneously in a social context (Menter et al., 2011).

In the analysis, responses were coded, categorised by key ideas and summarised (See Appendix F for example coding).

4.4.5.4 *Teacher Interview*

Semi-structured interviews were conducted with the co-teacher of my class before and after the intervention (see Appendix G for questions). As in the group interview, I wanted to enable flexible dialogue and allow the conversation to open in unexpected directions. The purpose of the first interview was to elicit another practitioner's perceptions regarding RQ1 and to generate ideas for potential interventions. The post-intervention interview was used to evaluate RQ2 and RQ3. In the analysis, responses were coded, categorised by key ideas and summarised.

4.5 ETHICAL CONSIDERATIONS

It was important to consider the potential effects of the research to ensure that the human dignity of all involved was preserved (Clark et al., 2021; Cohen et al., 2018). To do so, the guidelines set out by the British Educational Research Association (BERA) were followed throughout all stages of the investigation (BERA, 2024). Ethical approval for the project was sought and obtained from CUREC.

4.5.1 Consent

Formal consent was firstly gained from the head teacher at my school. As the students were deemed competent youth (Central University Research Ethics Committee, 2022), informed consent was obtained from them directly, with parents and guardians informed as a courtesy. To try and reduce the power asymmetry inherent in my relationship as the students' teacher, I left the room after providing the information sheet and time to read it and ask questions. Although Farrimond (2017) argues students will invariably feel a "subtle social pressure" to conform to teachers' expectations to participate, the intention in leaving the room was to give weight to the project as something separate from the normal course of lessons (p. 81). To emphasize the voluntary nature of the project, at each stage it was reiterated that the participants could withdraw from the study

without any repercussions on their studies or work (Cohen et al., 2018). However, other pressures, such as conforming to peer group norms, might have been at play (Brooks et al., 2014).

4.5.2 Power balance

In addition to the feeling of obligation my students may have felt to take part in the study, the power asymmetry in our relationship may have led them to providing responses aimed to please (Brooks et al., 2014; Cohen et al., 2018; Farrimond, 2017). To promote honesty, I emphasized the nature of the study to explore their authentic views. With an established student-teacher rapport, I strove for a balance in making the students feel at ease and that their contribution was important without placing expectations on them for any particular response (Cohen et al., 2018). I worked to remain as objective as possible, given my role as both teacher and researcher. As practitioner research cannot be value free, I constantly reflected on my own values and shared these as explicitly and transparently as possible (Menter et al., 2011; Tracy, 2010). My position as a practitioner-researcher must be acknowledged, however, and my inexperience as an interviewer may have influenced responses from the participants in unexpected ways (Menter et al., 2011).

4.5.3 Privacy

All efforts were made to keep personal information private. Pseudonyms were used and all identifying data was removed from documents, including gender-specific pronouns for the students. While the pseudonyms were used, the small size of the project and the use of qualitative data means using pseudonyms might not eliminate the possibilities of identification (Clark et al., 2021). Data was stored on my supervisor's One Drive and only kept for three years.

4.6 LIMITATIONS:

Several limitations to this study must be acknowledged.

Firstly, there is certain relevant information that could explain some of my findings that I cannot include for ethical reasons. This means some context is obscured from the reader, but confidentiality is maintained (Scheper-Hughes, 2000).

Secondly, collaborating with my co-teacher was more challenging than expected. Communication about the timing of different parts of the process lacked clarity at times. For example, an assessment happened in my colleague's lesson before I had communicated my intentions for the assessment review process. This resulted in a loss of potential data.

Finally, carrying out a study in a naturalistic setting within the constraints of a specific school context limits the applicability of the findings to a wider context (Cohen et al., 2018). While findings can potentially be transferred to other settings using "fuzzy generalisation" (Basse, 1999), readers will need to infer what is applicable in their own context (Denscombe, 2021).

5 FINDINGS AND DISCUSSION PART 1: DIAGNOSIS

5.1 RQ1 - WHAT ARE THE BARRIERS TO ANSWERING HIGHER ORDER QUESTIONS FOR MY STUDENTS?

Analysis was carried out across a range of data sources, which were used in conjunction with one another to elicit the barriers to answering higher order questions for my students. Question type, difficulty in question interpretation, difficulty with the language used in the question and an insufficient depth of knowledge emerged as barriers to the students in my class. Findings from the evaluation of this research question led to the development of the intervention, which follows in Section 6.

5.1.1 Question types

Questions that asked the students to make a judgement using evidence, deal with data, or to "suggest" formed a larger proportion of the questions which the students found moderately

difficult or more difficult, as seen in Figure 4. The questions which the students found less difficult were those that required them to recall, describe or explain.

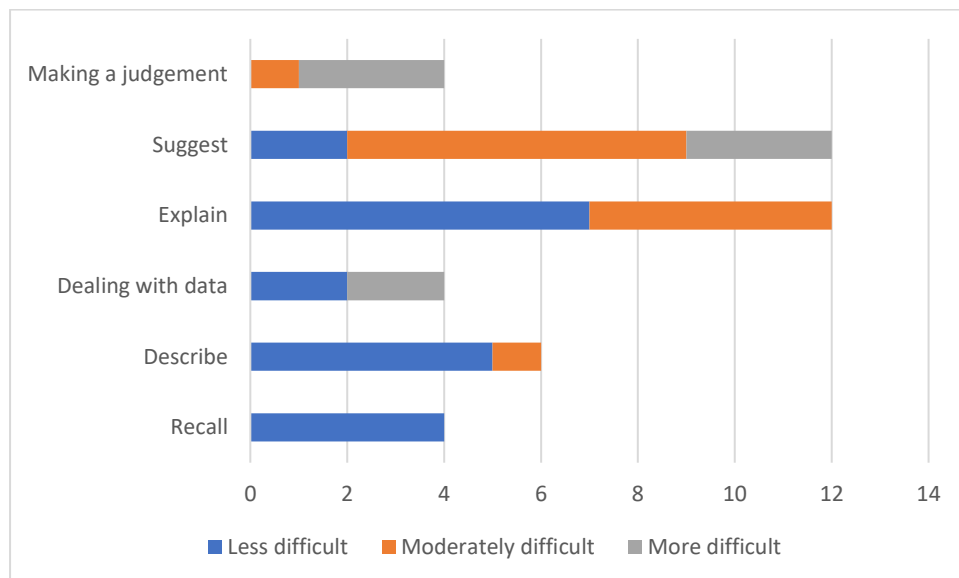


Figure 4: Distribution of different question types between the three categories of question difficulty from the first three assessments

Finding that students perform less well on questions requiring them to make a judgment, suggest a probable cause, or deal with data confirmed my expectations as a practitioner, having observed these (and other) students struggle with similar questions in the past. These question types are broadly in line with questions characterised by HOTS as categorized by the revised taxonomy (Anderson & Krathwohl, 2001; Das et al., 2022). To successfully answer these types of questions, students must go beyond retrieval of content knowledge to interpret new information in the context of their content knowledge, evaluate evidence, or provide a reasonable, possible suggestion, all of which can be considered HOTS (Bartlett, 1958; Lewis & Smith, 1993; Resnick, 1987; Sun et al., 2022).

The distribution of question type by command word into each of the three question difficulty categories is not clear cut, implying the influence of other factors on student success. For

example, “suggest” questions are found in all three categories. This indicates the command word is not the only determining factor of either question difficulty or student success (Jones et al., 2009). The topic and the wording of the question are two of the other factors that might influence the difficulty of the question for the student, providing a possible explanation for the range in attainment seen in similar question types (Das et al., 2022; Stanny, 2016).

When students were asked to identify the three most difficult questions during the first three assessments, there was a similar lack of a definitive pattern around question type and attainment for selected questions. While questions that asked students to “suggest how” or “suggest why” made up the biggest category of question chosen (see Figure 5), questions that require HOTS like, “Consider whether the graph supports the hypothesis” were chosen less frequently than questions that asked students to “describe” or “describe how”. Interestingly, in terms of attainment there were five instances out of 28 where students perceived the question to be challenging but attained full marks. In less than a third of the earmarked questions did students gain no marks. This finding contradicts expectations; however, parallels can be drawn with the findings from Jones et al. (2009). In their study of tertiary engineering students, they found no direct association between student performance and question level (Jones et al., 2009). Beyond a clear indication that “suggest” questions are perceived as more difficult, the lack of consistency in selected question type and varied attainment for those questions points to a more nuanced reason in determining question difficulty. There is also an implication that students’ metacognitive processes around monitoring are inaccurate.

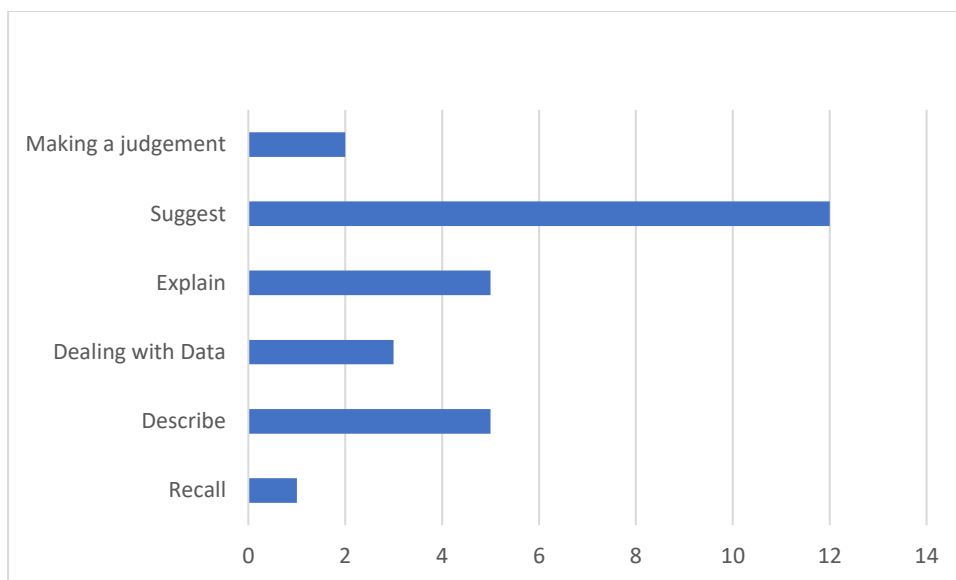


Figure 5: Distribution of most difficult questions as selected by students during the first three assessments

One factor that seems to increase the challenge for students is the fact that questions demanding HOTS often include unseen information, particularly unfamiliar data and graphs. In the pre-intervention interview, my co-teacher identified the most challenging questions for students as those that require an evaluation of a conclusion or include unfamiliar information or data:

Some of them just panic. And then when [the question] is built around data, for example graphs, and they have to extrapolate or infer something from a study that is clearly related to something that they have done in lessons, but I think they find it hard sometimes to find those links.

His comments point to the complexity of answering such questions. A student must undertake several cognitive steps before crafting a response. First, a student must understand and interpret the unfamiliar data, graph or information, then infer meaning, and lastly link it to their content knowledge. Students may lack the HOTS to understand and break down the problem into separate, solvable steps as suggested by Pólya (1957). Without the necessary HOTS to address each part of the question, each step required might become a barrier. Thus, it is not surprising some students “panic” when faced with unknowns and therefore find it hard to draw on their own knowledge.

5.1.2 Difficulty in question interpretation

Difficulty with question interpretation commonly emerged when discussing question difficulty. When identifying the three most difficult questions during an assessment, Student B commented there was, “Lots going on with the graph. Bit confusing wording.” This comment draws attention to their difficulty in interpreting the graph presented in the question. For them to state there is “lots going on” with the graph implies they find it challenging to break down the data for interpretation. Similarly, Student C states they are “unsure what the question wants”, indicating they are unable to understand the full requirements of the question in context. If students feel unable to understand the demands of the question, this will prove a major barrier to answering questions.

Finding that students’ perceived understanding is a barrier was replicated in the student responses to the post-assessment reflection tasks. Analysis of data from the post-assessment coding shows most reasons for not attaining marks were attributed to all types of questions but were not distributed evenly, as seen in Figure 6. For the most difficult and moderately difficult questions, students attributed a lack of understanding of the question (Q) more frequently than they did for less difficult questions. A lack of understanding of the topic (U) was also more likely to be ascribed to the most difficult questions.

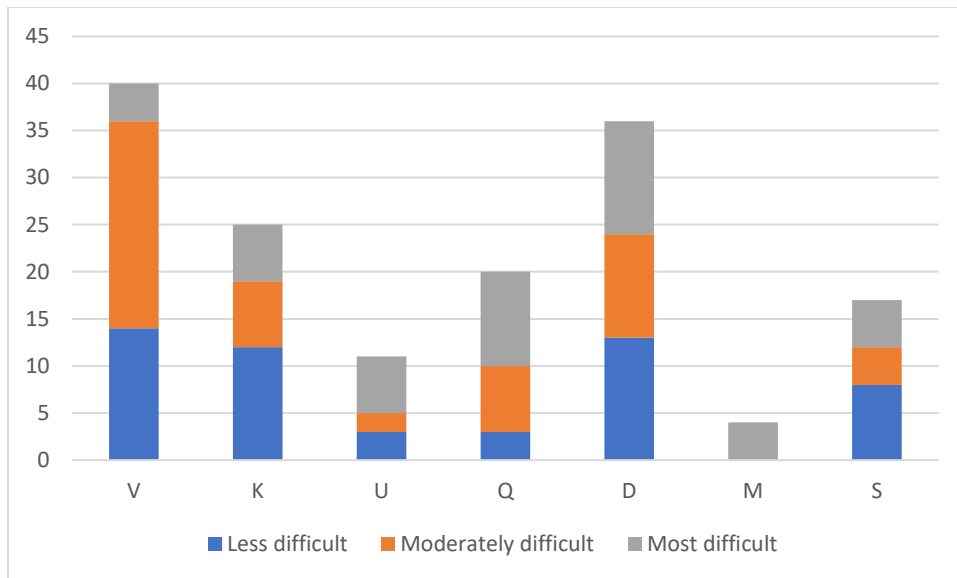


Figure 6: Distribution of 153 total reasons given for not gaining full marks. Collected from 43 questions.

Comments in the questionnaire mirror the data in Figure 6 and provide more context. In response to the questions, “What did you find challenging about this set of questions, as a whole?” and “What are the reasons you lost marks? Put the biggest reason first,” “not understanding the question” was one of main reasons mentioned by the students. Variations on this comment included references to “misreading the question”, not reading the question properly, and “a lack of time to read [the] question and understand it”. These comments imply either that students did not have enough time or did not take the time to fully comprehend the question. My colleague commented in his interview that students do not always spend enough time looking at the data before responding to the question. Time pressure associated with an assessment may also impact students’ abilities to understand the demands of the question and complete each cognitive step. It must also be acknowledged the comments about misreading the question could also implicate reading comprehension as a barrier, which will be explored later.

In our interview, my colleague agreed that understanding the question can be challenging, particularly when the scenarios presented in the question are unfamiliar. He said students are trying to do two things:

They're trying to decipher what the question is actually trying to ask them, and then they've got a selection problem. Those better students have to decipher the question still, but once they've done that, the knowledge is there for them to kind of just fit it in, a bit like pieces of a jigsaw.

Like Pólya (1957), my colleague acknowledges understanding the problem is the first step to answering the question. As noted previously, the possibility exists that students lack the necessary skills to break down the information into solvable steps or struggle to transfer skills applied to previous problems to the novel one (Orton, 2004). Their lack of understanding of the question prevents them from constructing a response which is fully aligned with the demands of the question and may lead them to randomly guess (Muijs & Reynolds, 2018) or avoid the question altogether.

5.1.3 Difficulty with the language of the question

The wording of the question came up in different data sources as a reason for question difficulty. This was not surprising, as understanding the domain-specific language used in the question is an aspect of understanding the question. In Figure 6, above, although “understanding vocabulary” was not specified as a reason, both a lack of understanding of the question (Q) and a lack of understanding (U) may have been chosen by students when they found the wording difficult. One student explains how a particular question was difficult because of the link between unfamiliar academic language and understanding of the question. All participants pointed to wording of the question as a reason for difficulty either during or following the assessment on cell membranes, with most citing a specific question. The question, “Describe the mechanism for the absorption of amino acids in the ileum,” required only a demonstration of knowledge and understanding of the topic and not any HOTS. Thus, it was surprising it caused difficulty in the

group. In their explanation of why this question was difficult, both Student B and C point to the word “mechanism” as the root of difficulty. “Mechanism” would be considered a Tier II word (Beck et al., 2002). As a non-content related word, it is difficult to define using the context of the question, so it is not recognised by the student. This highlights what Snow (2010) found about the barrier academic language can pose to understanding the text.

For most students, however, comprehension of text was not deemed a major barrier overall. In the interviews carried out to explore reading comprehension, all students demonstrated they were able to retrieve information from the text and infer the meaning of a scientific or technical word in context. Speed of retrieval from the text varied between students. None of the students seemed confused by any of the language used or had coping mechanisms for unfamiliar terms. Student B said they “blanked” out unfamiliar terms, such as unrecognized names of drugs, and treated them like “a little box”. With their comment, Student B demonstrates an ability to decipher the text without knowledge of those words. It was concluded that for all students except one, the retrieval of information and inference did not seem to be a major barrier to understanding.

Notably, one student found the reading comprehension more challenging. They re-read the text several times before they could retrieve the information needed to answer the question in the interview. Often, this student noted the wordiness of questions or the length of questions as reasons for the difficulty they experienced. This could be explained by the student’s processing speed, but also by their cognitive load. (Holman & Yeomans, 2018; Sweller, 2010). The differences noted here highlight the necessity to keep individual needs in mind, as not every student will experience the same barriers to answering a question.

5.1.4 Insufficient depth of knowledge, including a lack of biological vocabulary

Insufficient content knowledge was another common thread to emerge from my analysis of several sources of data. The student comments below point to lack of content knowledge with statements around their uncertainty of their grasp of absorption as a topic:

“Mechanism? What is it? Also need to revisit absorption. Unsure what question wants.”
Student C

“Wasn’t sure which type of transport amino acids are absorbed by.” Student A

In Figure 6 above, a lack of knowledge is the third most frequent reason given for losing marks, particularly for the less difficult questions. Students elaborated on this in the questionnaire, where they often mention a “lack of knowledge” and a “lack of revision” for reasons they lost marks, or that “more revision is needed” indicating that they attribute their lack of success to insufficient depth of knowledge or an inability to retrieve that knowledge.

When the students were questioned about the content knowledge needed to answer the examination questions in the reading comprehension interviews, their answers were less confident and required further probing, compared to questions about retrieving information from the text. This was particularly true for students who are typically lower attaining. Given that lower-attaining students often have less domain knowledge, and that existing domain knowledge has a substantial impact on understanding and interpreting text, this finding is unsurprising (O’Reilly & McNamara, 2007).

Part of this perceived insufficient depth of knowledge is a lack of knowledge about the domain-specific vocabulary. Students struggle to use the vocabulary expected from an A-Level biologist. In Figure 6 above, not using the correct vocabulary (V) is the most frequent reason given for losing marks, particularly for less difficult and moderately difficult questions. Not including enough detail in the written responses (D) is the second most frequently given reason for losing marks and is distributed equally across question difficulty. Interestingly, Student A consistently

attributes V and D to their lower attainment more than any other reasons. Student A confirms these choices through their comments in the questionnaire by referring to vocabulary or not including enough detail. Choosing both V and D points to a lack of depth of domain-specific knowledge when constructing a response to the question. Student A finds it difficult to articulate their understanding of the concepts using domain specific language, which results in gaining fewer marks.

Another aspect of this insufficient depth of knowledge is an inability to differentiate between concepts within a topic and select the appropriate ideas. In the section above my colleague's comment implies students must select the appropriate content knowledge after they determine the requirements of the question. Student B highlights this after the assessment on the ecology topic by responding in the questionnaire that it was challenging "not understanding what aspect of the topic the question wants you to write about." This comment implies a lack of understanding of the question, brought about by insufficient content knowledge. She goes on to state she lost marks because "I couldn't completely differentiate between different sub-topics. They all interlink and merge in my head." In my experience as a practitioner, students often cite that they struggle to apply their knowledge to the question, as noted earlier by O'Reilly and McNamara, 2007). My colleague also mentions this in his interview. He referred to knowledge as 'pieces of a jigsaw'. Using this analogy, an insufficient grasp of the content knowledge and vocabulary is like being unable to find and fit the correct piece into the jigsaw from amongst the myriad similar pieces spread on a table.

6 DEVELOPMENT OF INTERVENTION

From my initial data collection and analysis, several potential foci emerged for an intervention. The findings indicated question type plays a role in perceived difficulty. Attainment was often

lower in response to questions requiring HOTS, but the data revealed that the barriers around these higher order questions were complex and nuanced. Understanding the demands of the question by deciphering the language of the question and any unseen information also play a role. Additionally, selecting the appropriate content knowledge (if present) proved challenging. The multiple barriers faced when answering higher order questions for these students suggested that a multi-faceted intervention focusing on strategies used to approach questions might be useful.

Student questionnaire responses about how they might approach questions differently in the future show that they are cognisant that their strategies require improvement:

“Re-read the question”

“Re-read answers and questions to refine my points in a suitable order.”

“Take time and stay on track of what the question is asking.”

“Slow down and highlight key words in questions. Interpret what the question asks only from these key samples.”

“Try and analyse it more to find out more detail on what it wants me to write about when the questions aren’t as obvious, looking at the questions before and after might help me do this too.”

There are multiple comments here highlighting the students’ varying levels of awareness that they must grapple with the question on a deeper level. Some refer simply to re-visiting the question, without any indication of what they might gain from doing that. Others recognise a need to identify key aspects of the question that provide the context for the demands of that question.

Two students demonstrate an awareness of using a metacognitive monitoring process to look “at the questions before and after” or “stay on track of the question”. Alternately, it could be argued that “staying on track of the question” refers to a student’s struggle with keeping all aspects of the question in their working memory. Collectively, these comments imply the students might benefit from improving their cognitive and metacognitive knowledge and strategies.

An intervention was developed to explicitly teach some cognitive and metacognitive strategies to use when approaching higher order questions, as the literature suggests direct

instruction is the most effective (Holman &Yeomans, 2018; Muijs & Reynolds, 2018; Pólya, 1957). A think-aloud approach, inspired by those I came across in the literature, was chosen to model strategies selected during collaborative discussions with my co-teacher (Allen & Hancock, 2008; Dabarera et al., 2014; Holman & Yeomans, 2018; Jackson, 2016; Muijs & Reynolds, 2018; Rosenshine & Meister, 1992). During the initial intervention lesson, students were given a cue card with the strategies outlined in Table 8. Collectively, the students referred to the strategies as “SKLIS ME”. My colleague suggested using a cue card to introduce the strategies, an idea supported by Rosenshine & Meister (1992).

Strategy	Action
S - Scan	Read and scan the question – what is the general idea/topic?
K – Key words/data	What key terms are being used? Key data?
L – Link	Make links between the question and what you already know. Make some notes/draw pics
I - Instructions	What is/are the command word(s)? Are there other instructions? How many marks?
S – Summarise	In your own words, summarise what the question is asking you to do.
M – Monitor	Is what you are doing working? Are you answering the question/including the correct number of marks, etc.
E – Evaluate	What worked? What didn't?

Table 8: Strategies given to students (on cue card) and modelled by teacher

The use of these strategies was then modelled by the teacher, using a question that included unseen data and selected to provide desirable difficulty (Bjork & Bjork, 2014). Following this, the teacher led the students through a scaffolded question (Rosenshine & Meister, 1992). (See Appendix H for an example of a question annotated during a think-aloud modelling). Lastly, the students worked in pairs, taking turns to think-aloud the strategies with each other on subsequent questions. In an attempt to execute the intervention strategies in a similar way, my colleague observed the introductory intervention lesson I did with the class. The intervention took place over a five-week period, mostly in my lessons with the group, but also in several lessons with my colleague.

7 FINDINGS AND DISCUSSION – PART 2: POST-INTERVENTION

7.1 RQ2 – WHAT IMPACT DOES AN INTERVENTION HAVE IN ADDRESSING THE BARRIER TO IMPROVE SUCCESS IN ANSWERING HIGHER ORDER QUESTIONS?

7.1.1 Confidence

Overall, confidence of the cohort increased in their ability to answer more challenging questions following the intervention. In the post-intervention questionnaire, each of the five students answered positively when asked how their confidence in answering more challenging questions had changed since January. They used phrases like “more confident”, “better”, and “improved”, but four out of the five students qualify this by saying confidence has improved only “a little bit” or their approach to harder questions “still needs work”. The way in which they temper the gains they have made highlights their perception of a need for more practice in answering challenging questions to develop their skills further. It also suggests their metacognitive awareness of their own abilities may have increased, such that their self-efficacy assessment is more accurate (Linnenbrink & Pintrich, 2003).

Familiarity with the more challenging question types seems to play a role in confidence. In the group interview, Student E remarked that regularly being given difficult questions in class prepares them for an exam:

Because now when we get into an exam, for example that exam we just did, and [the questions] are not all six-mark or four-mark “suggest” or “evaluate”. And it prepares you. It’s like, well, I can actually answer this. And then that makes me feel good.”

A similar thought was shared by Student C who remarked, “Suggest questions are not scary any more”. This comment implies increased exposure to difficult questions increases self-efficacy in two ways. Firstly, it reduces the anxiety that can be caused by challenging questions by normalising a higher level of difficulty (Palmer, 2005). Thus, students are less likely to link anxiety to inability.

Secondly, as the student experiences their own success and the success of others, self-efficacy is increased through both mastery and vicarious experiences (Bandura, 1977; Palmer, 2005).

My colleague agreed confidence has grown across the cohort and that exposure to questions is important. In the post-intervention interview, he echoed the sentiments of the students that there was less panic when faced with difficult questions, particularly for lower-attainers. He expands on how the students have benefited:

I think if you put a question in front of them six months ago that was a big block of text with a table or maybe a graph in it, their natural response would be to recoil...but I think if you put it down now in front of Student E, because [they have] a scaffold and because [they have] seen that it works when [they've] applied it...[Student E] knows that actually they don't need to panic.

His remarks point not only to familiarity with difficult questions as playing a significant role in increasing self-efficacy, but also links the scaffolded thinking strategies to self-efficacy by providing a structure the students can use to approach these questions.

7.1.2 Structured approach shifts thinking

The idea the thinking strategies provide a structured way to approach questions, and thus make students feel more confident was shared by the students in both the post-intervention interview and the post-intervention questionnaire. When asked how participating in the study has changed how they think about and respond to questions, Student E notes,

I think at the beginning of A-Levels you just get freaked out by questions because they're so long and so much information. But now I actually don't mind them as much because I actually find if I use the method, it actually gives me an answer sometimes.

Student E confirms what my colleague said above: more difficult questions, particularly those that are text heavy, previously caused them to panic. The strategies introduced through the intervention allowed students to approach longer questions so they can craft a response with more confidence. Using a structured method to break down the longer questions allowed this student to access the ideas in the question that had been previously inaccessible. The intervention

enabled this student to use more metacognitive strategies and HOTS in answering the question, potentially by reducing the cognitive load through a problem-solving heuristic (Pólya, 1957; Sweller, 2010).

In response to the questionnaire question about what had the biggest impact on confidence in answering questions, many linked the thinking strategies and modelling to an improved approach to questions. Both students below refer directly to the cognitive strategies used in the intervention as having a positive impact on their confidence since the beginning of the project:

Going through thought processes for going through difficult questions, looking at examples of how to answer questions to make sure you have the right points. Student A

The SKLIS ME strategy and the help that I have gotten with evaluate and suggest questions. Student C

In their comments, they acknowledge a shift and development in their thinking when approaching questions. While the thinking processes have not been measured directly, in the post-intervention interview Student C describes how they now take more time to process the information presented in a question, using the strategies as a starting point to analyse it:

Time taken to think has definitely been extended for me. And I approach stuff from different angles now. I approach it more from the sense of the keywords and in the scan of the question I do, rather than the instructions. I approach it from looking at the question...after scanning it and then answering and then kind of making sure that I have summarised the instructions because before I used to look at the instructions and then follow what I thought they wanted me to do, and I honestly would read about a quarter of the question.

In the response, this student discusses the merits of using the different strategies introduced in the intervention, such as scanning the question and identifying keywords at the start of a question, demonstrating that they have engaged with the process. There is a sense that the student is decoding the question methodically, followed by an evaluation of whether the answer follows the instructions. This student's metacognition and problem solving-skills have been positively

impacted by the explicit teaching and modelling of these skills (Hembree, 1992; Holman & Yeomans, 2018; Muijs & Reynolds, 2018).

The impact the intervention had on extending the time this student, and others, take to read and decipher the question is an intriguing finding. From my own practitioner experience, I have observed students often skim-read questions, without fully processing the information. In the post-intervention interview with my colleague, he agrees:

Students have a tendency to jump straight to [answering] the question, with the lines to fill, without looking at the graph or reading the information.”

He feels they are less likely to do so now after the intervention. If students are taking more time to decipher the demands of the question before answering, this may eventually lead to more success in answering the questions as students gain a better understanding of the questions.

7.1.3 Attainment

There was no discernible change in mean cohort attainment over the period of the research project. Attainment in the sequence of tests fluctuated for all individuals, with some performances varying greatly (see Figure 7). Students A and C both demonstrated a small increase in attainment, which could potentially be attributed to the intervention, however, the other three students' attainment either decreases, stays the same or fluctuates too widely to draw a conclusion. The performance of the three students who do not yet see a sustained increase in attainment could be viewed through the lens of several contextual reasons. For ethical reasons around privacy and confidentiality, this is not possible.

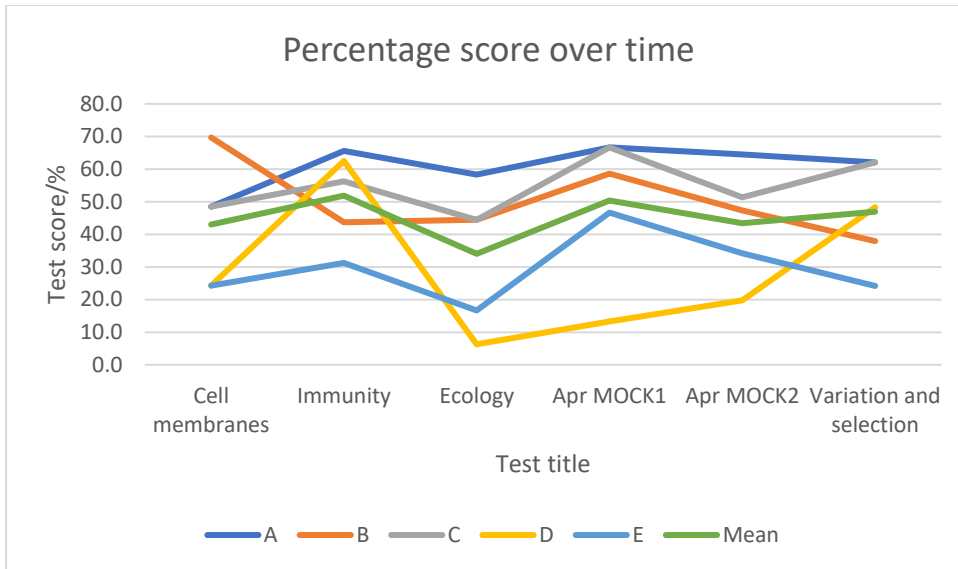


Figure 7: Individual and mean test scores over subsequent assessments which took place between January and May

There was not a perceptible shift in which questions the students found most challenging either (see Figures 8 and 9). While some potential patterns could be elicited, the small number of questions in the post-intervention assessment makes the data less comparable.

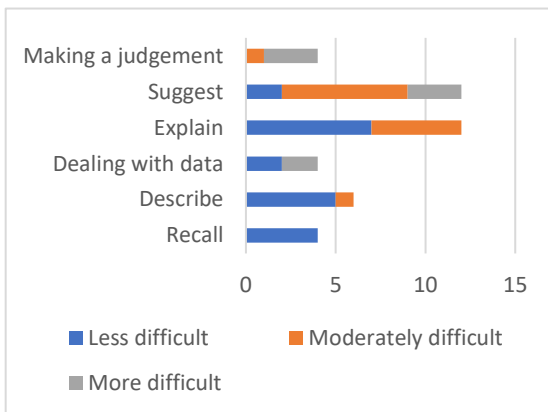


Figure 8: Distribution of question type between question difficulty categories, pre-intervention

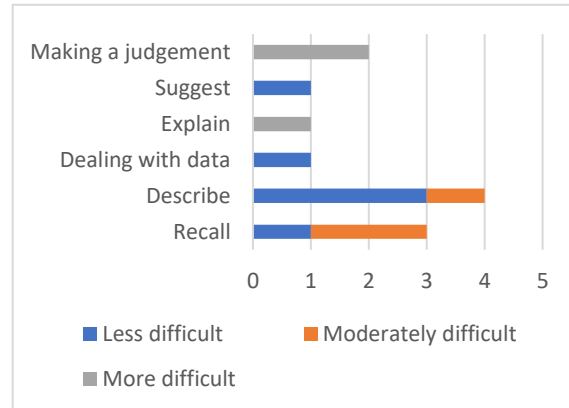


Figure 9: Distribution of question type between question difficulty categories, post-intervention

Attainment might not yet have been impacted by the intervention, as a number of barriers may still need to be addressed. Students perceive practice and revision as playing an important role in their ability to answer more challenging questions. In the post-intervention questionnaire, Student B touches on both ideas:

It's gotten better but I need to consolidate my knowledge and links between different topics to make them easier.

Practice has definitely increased my ability to approach them but as the course links together more and I know all the content to a similar standard, I think I'll be able to practise my answering techniques more effectively and gain higher marks.

While Student B's confidence has improved, they still feel limited by their knowledge when faced with difficult questions (Willingham, 2021; Zoller, 1993). Rather than referring directly to the cognitive strategies used as having the biggest impact on their confidence, Student B again mentions needing to know the content. There is a recognition they employ "answering techniques", although it is not clear whether these are the modelled techniques or others. Given that this student's work often contains annotations, similar to those used in the modelled strategies, it can be inferred that this student feels these techniques require practice and deeper understanding of the content to develop further.

Linking to the effect of content knowledge, the topics also impact confidence in answering difficult questions. Cohort confidence in the topic varies from test to test. This is not unexpected, as student preparation for the test, difficulty of the topic, and student comfort with a topic will also be different throughout. While confidence about topics, as measured using a Likert scale in the post-assessment questionnaires, generally increases after taking and reviewing the test, this trend is slightly reversed in the April Mock (see Figure 10). This reversal could be due to the wider selection of topics assessed. Students did not know which exact topics from a broader selection would be chosen, so they might not have reviewed their knowledge of all topics as fully.

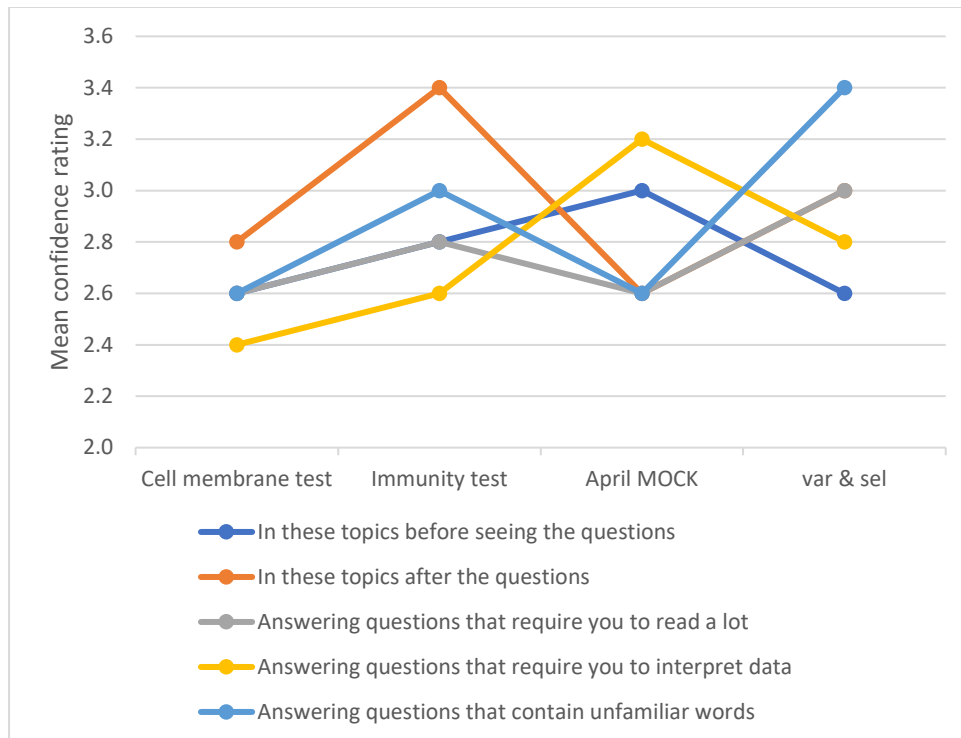


Figure 10: Change in mean confidence rating over a series of assessments between January and May

There is a general upward trend in confidence in response to all statements about more challenging questions, including questions requiring more reading, data interpretation or that contain unfamiliar words. The differences, however, are very small and fluctuate over time. The small variation in mean scores could be attributed to the nature of Likert scales and human propensity to pick more middle values (Cohen et al., 2018), as the scores do not reflect comments fully.

7.2 RQ3 – WHAT ARE STUDENT AND TEACHER VIEWS ABOUT THE PROCESS OF DIAGNOSIS AND INTERVENTION?

7.2.1 Test review using key

One unexpected finding from the diagnostic phase is that both students and co-teacher viewed the process of reviewing assessments using a key as helpful. Three reasons for this emerged from the data.

The first idea to emerge from the interviews is that through the process of coding mistakes, students perceive mistakes as more manageable. When asked about the usefulness of reviewing test performance and coding them afterwards:

It's taken the focus away a little bit from just a general sense in their mind of 'I failed this exam where I've not got the grade I was looking for.' Now, their focus is much more on 'why have I not done as well as I can?' The result hasn't changed but, actually, for them the formative nature of the assessment is really coming through. Teacher

It makes mistakes feel more manageable. Like you can improve. And where to improve as well. Student B

An important shift in student perceptions is identified, from one of failure and finality to one of learning and improvement. My colleague reflects on the formative nature of the assessment being emphasized, which is echoed by the student. The feedback allows the student to reflect on their performance and determine the direction of future learning. The student also believes they can improve, demonstrating greater self-efficacy.

The idea that the assessment review process helps shape future learning and revision was the second idea that emerged. Student B mentioned it above, when they spoke about 'where to improve', and others agree:

I can focus my revision. For example, if I've got a bunch of Ks...(knowledge), I just don't know the topic. Just go back over and...revise. If it's a bunch of Qs, so I don't understand the question, then I can, like, focus my revision...I'll have a look through some four-markers in the practice paper books and try to nail them. Student E

It gets me to figure out why exactly I've been losing the marks and it helps to, you know, specifically what I can target in revision and stuff. Student A

I get to plan. Normally, when I review my tests, I'm not a big one for planning stuff. I like to just mark and I go, 'That was wrong,' but using your approach to actually force me to think about what I need to do, which allows me to like kind of plan out. Student C

There is an increase in the student's self-knowledge from the evaluation of their assessment, which feeds back into their knowledge of their metacognitive processes (Muijs & Bokhove, 2020).

Their evaluation of their work allows them to plan the next steps of their learning. Rather than

blindly moving from one assessment to the next, there is conscious reflection involved. The idea of the role of metacognitive processes is addressed by my colleague:

...now they are able to very quickly understand why they didn't get the marks, and because of that, this idea of a bit of metacognition, I guess that they can start to understand what they would do to put them right because they have experience of lots of mistakes in the past and they can categorise them in their mind.

The increased metacognition in understanding mistakes improves self-efficacy so they believe they could improve in future assessments.

The third idea that emerged from the interviews was that the students started to view the coding process as part of their learning process in biology and, for some, in other subjects. While none had used this specific process for reviewing assessments before, all agreed in the interview they would use this strategy to review future tests. Two of the students remarked they felt the codes could be refined further for personal use. One student had already used an additional code in the post-intervention test review to denote a mistake associated with a graph. One of the more reflective participants provided evidence that the approach to test review can be transferred to other subjects in the notes made before the group interview:

I use the reviews in my chemistry to help increase confidence and understand and visualise what went wrong.

The comment above demonstrates that the process of identifying explanations for perceived failure lessens the sense of failure. As in the finding in RQ2, the formative, metacognitive nature of the process seems to boost self-efficacy. While this potentially contradicts what Yerdelen-Damar & Pesman (2013) found in relation to the mediating effect of self-efficacy on metacognition, it aligns with Sun et al.'s (2022) finding that by positively affecting scientific reasoning, metacognition enhances self-efficacy. My findings here support the suggestion that the relationship between metacognition and self-efficacy is both important and complex.

7.2.2 Perceptions of using the intervention strategies

A key finding from the group interview is that all students perceive the list of intervention strategies as useful when approaching challenging questions.

I think it's quite useful 'cause like, sometimes I might not know how to approach a difficult question and it's helpful to know where to start and how to go about answering it. Student A

I tend to get flustered by questions. All my thoughts like jumbled up, so when I have a structure I can follow, it might break it down a little so I can answer it. Student D

The comments above note the structured nature of the strategies gives the students a starting point, particularly when there is a feeling of under-confidence in relation to a difficult question. Student D's comment reveals they feel the intervention has provided a way into questions previously viewed as inaccessible. This finding is particularly interesting as this student has a tendency to leave questions un-answered if their perceived self-efficacy for that task is low. The strategies provide a scaffold for the student to find some success. This leads to greater self-efficacy, and so the student is more likely to engage with the question (Linnenbrink & Pintrich, 2003).

The students perceive that using the strategies has enabled a deeper level of cognitive engagement.

I think it makes a structure to make sure that you focus completely on the question and don't rush into the answer without knowing exactly what you're going to say, and then that's aligned with the question, what it's asking. Student B

I find myself noting in the margin and it links it together a bit more because sometimes the endings are quite far reached between the question and the knowledge. It helps my thinking, and it helps my focus throughout the question. And I can manage the question more easily. Student E

Both comments allude to the usefulness of the strategies in focusing the students' thought processes and for providing a heuristic to solve the "problem" of the question. Student B's comments point to the idea of engaging in metacognitive monitoring and planning. The annotations Student E makes provide an example of how representing the problem can lead to a

better understanding of it, enabling links to be drawn between their prior knowledge and the new information (Hembree, 1992; Pólya, 1957). The idea about “focus” while answering the question, suggests again that such strategies help reduce cognitive load, allowing more cognition around the question. By inference, the metacognition of the students has increased as they can reflect on which cognitive strategies are proving most useful.

A major drawback of using the strategies perceived by the students is it requires too much time during an assessment. As these are new skills and the process is very detailed, this is not entirely surprising. Indeed, one of the positive impacts of the intervention noted above is that it has encouraged a closer reading of text than the students would have done otherwise. As the students continue to practise these skills and their exposure to questions increases, it is not unreasonable to suggest their skills will develop. This will allow them to decipher questions more quickly, reducing the time pressure felt during assessments. Barriers may remain, however, that need to be addressed before the above happens.

7.2.3 Perception of teacher modelling

Students’ perceptions of teacher modelling were generally positive, with a few caveats.

Students found the modelling process useful at the beginning of the intervention, for the start of a topic and for difficult questions, as the model they observed can then be applied to similar questions:

Quite useful because sometimes I might not know how to approach a difficult question and it’s helpful to know where to start and how to go about answering it. Student A

I know how to approach similar questions, so now it breaks it down more so I can answer the question. Student D

I find it sort of helpful the first time...It might be helpful if you’re kind of in a blank state that you don’t know how to answer the question, so for a new topic, that’s very useful. Student C

Students felt that the modelling should be phased out and not overused, however,

If you do it too often, then you're kind of forcing the way of thinking onto people. Just going through kind of a test, when everyone's already done this, they know it. It's sometimes not that helpful because you think in different ways. So, you've obviously taken on board what they explained the first time and then normally people would do it like that, but it would be in a slightly different way. Student C

The statement above is an interesting view on expert thinking, where there is a rejection of the thinking of the expert in the effort to try to develop one's own expert-like thinking (Rosenshine & Meister, 1992). While the student takes on board the advice at the outset when they feel very inexperienced, this changes after they have attempted and answered the question themselves. What is striking is the lack of recognition that the student's thought process might not lead them to an accurate answer, even if they had used the strategies. When probed, student C elaborates to the agreement of others in the room:

Then it's not a point of you modelling your thinking to me. It might be better if I model my thinking to you and then you could say, 'That's where your thinking needs to change.' Because otherwise I'm comparing how I thought about it and how you thought about it... So, it's just those small tweaks, rather than me having to redo my whole kind of thought process.

Here there is a *recognition* the student's thought process is developing, using the given strategies.

There is a desire to have help in developing those thought processes further through feedback, rather than through repeated modelling. Student E echoes this thought:

I liked it, but I definitely think it was as it approached at the end. Like if I get used to you guys going through, like actually I need to read the question. Have a think about it before you go through. Otherwise I kind of find myself. Ohh [leaning back], they're just doing the work for me. I'm just listening. That's just cause I've got a short attention span. If I look at the question first, oh, actually, how do I do this? And then I'll listen then I'll get 'well, actually they did that really clever thing. They'll find the links after I tried to.

The students feel they were ready for the next step in the scaffolding process and identify the need to attempt questions themselves before hearing teacher thoughts. Formative feedback on the attempted procedure is desired and is an important part of the scaffolding process (Bohn et al., 2004; Rosenshine & Meister, 1992). While the scope of this intervention did not include feedback, it would make a logical next step in a teaching progression.

7.2.4 Conscious teacher planning

My co-teacher found the intervention practice useful and thought-provoking. He perceived his involvement in the project as a way of formalising aspects of his practice that were previously less deliberate. He noted that while the intervention is “not an epiphany,” it provided a structure for dissecting examination questions as a class, thus making the information more accessible to the students. His perception of how implementing the intervention has impacted his teaching mirrors many of the perceptions of the students.

Firstly, he feels he made a more conscious decision about the types of questions he selected when modelling his thinking:

So, before I would just select a variety of different questions, and I couldn't tell you necessarily why I chose them other than they looked good without kind of thinking about it more. Now I'm looking for certain types of questions where you can apply the SKLIS ME model to it.

The types of questions he now chooses are those that require data interpretation. While he feels he previously tried to model his thinking, he now does this in a much more conscious, structured way, similar to what we were expecting of students.

Like the students, my co-teacher felt that the next step would be to move on from more formal scaffolded “I do, we do, you do” to where the students are more flexible in their approach and less formulaic and mechanical in the application of the scaffolded strategy instruction. The aim would be to get the students to develop a more fluid use of the strategies so “they can apply what they know needs to happen, but in different contexts.”

A last finding from the post-intervention interview with my colleague was that he found the process “thought-provoking:”

What must it be like from the point of view of the students when we give them really hard questions that they repeatedly fail on without improvement. Seemed like hard work to get any improvement.

The perception of both the teacher and the students is that the modelling and the practised strategies have provided an access point to questions that could otherwise be dishearteningly inaccessible. My observation, with which my co-teacher agrees, is the humbling nature of thinking on the spot and making mistakes in front of the students. Our hope is that modelling imperfection makes success feel more attainable, thus reassuring the students. As my colleague pointed out, there is a message that 'we all struggle sometimes to make sense of what's on the page.' The perception is that the strategies have done some work in increasing self-efficacy for the students, potentially through vicariously experiencing a teacher's cognitive struggle (Palmer, 2005).

8 CONCLUSIONS AND IMPLICATIONS

8.1 SUMMARY OF FINDINGS

This research aimed to diagnose and address the underlying mechanism for low attainment when responding to higher order questions in my students. In this section, I will summarise the evidence found for each of the research questions, followed by a consideration of the implications of my findings.

8.1.1 Barriers for this cohort to answering higher order questions

Overall, the findings in the diagnostic phase highlighted the barriers my students face when answering higher order questions are complex and nuanced. Broadly, attainment was lower on questions that required HOTS compared to those that simply required recall, knowledge, or understanding. Students particularly struggled with questions that required them to make a judgement or suggest a probable cause. The data imply, however, that question type is not the only factor influencing question difficulty or student success.

Difficulty interpreting and understanding the demands of the question was another factor found to be an influential barrier to success. Perceptions of question difficulty increased

particularly with the inclusion of novel information, such as data. Additionally, the findings suggested the students struggled, at times, to understand the complex academic language of the question (Snow, 2010). Both the unfamiliar information and complex vocabulary prevented them from applying the appropriate heuristic to the problem, especially within the time limits associated with assessments (Muijs & Reynolds, 2018; Orton, 2004).

Thirdly, insufficient depth of domain-specific knowledge, including a lack of biological terminology, was found to be a barrier to answering higher order questions in my students. The superficial nature of students' content knowledge affected their ability to differentiate between similar concepts. Thus, students found it difficult to select and use appropriate knowledge and vocabulary in their responses. While my intervention included a strategy to link existing knowledge to the question, it did not propose to address knowledge building directly. Higher order retrieval practice might be another avenue to explore in future practice or research to address this deficiency (Agarwal, 2019).

8.1.2 Impact of the Intervention

The intervention had a positive impact on the confidence of the cohort in responding to more challenging questions. The familiarity gained through increased exposure enhanced self-efficacy by reducing the stress normally associated with these types of questions. Increased mastery experiences and the vicarious experiences provided by think-aloud modelling by the teacher and peers also increased self-efficacy.

The modelled strategies provided a helpful structure for students when approaching more challenging questions. Modelling enabled the cognitive and metacognitive strategies to increase (Muijs & Bokhove, 2020; Schraw & Dennison, 1994). The strategies also may have supported a reduction of cognitive load, as the question was broken down into smaller, more manageable steps

(Sweller, 2010). The structure provided by the explicitly taught strategies also helped shift and develop the thinking of some students, possibly providing evidence for increased HOTS.

A lack of discernable impact on the attainment of the cohort suggests that other barriers need addressing before cognitive and metacognitive strategies can be developed further. Domain-specific knowledge is cited by the students as one potential barrier (e.g. Willingham, 2021). Since topics are successionaly taught and assessed in A-Level Biology, students never have the opportunity to reach mastery of the content in lessons before moving on to the next topic. My experience is that the school structure and wider educational structure do not enable this. If students are to gain more secure content knowledge, they are expected to do so outside of lessons.

8.1.3 Student and Teacher Views

Both the students and my co-teacher perceived the diagnostic process of using codes to review assessment performance as helpful. Several reasons for this emerged. Firstly, the reflective process made mistakes feel more manageable. Secondly, the formative feedback was recognised as being helpful in shaping future learning. The skills the students gained in evaluating their performance and planning ways to address shortfalls increased metacognition. Lastly, there was a perceived increase in self-efficacy, potentially linked to increased metacognition, supporting findings from Sun et al. (2022). The relationship between self-efficacy and metacognition was not explicitly explored in this project, however.

Students also perceived the cognitive strategies (SKLIS ME) introduced during the intervention as useful. The strategies seem to provide a scaffold or heuristic that the students can use when faced with an unfamiliar problem. This enables students to find some success, which then leads to greater self-efficacy. The potential reduction in cognitive load mentioned above, enabled further thinking about other aspects of the question. As their thinking skills continue to develop, using the

strategies will become more automatic, and take less time. Lastly, the students indirectly recognise their metacognitive skills have improved. Potentially, their learning could be improved further through additional explicit instruction of metacognitive strategies (Muijs & Bokhove, 2020).

Although the teacher think-aloud modelling was perceived as a useful starting point for developing the cognitive strategies, the students felt ready for less scaffolding. There was a sense some students perceived their thinking skills had developed. They desired formative feedback to improve further. My co-teacher agreed the class was ready for the removal of formulaic scaffolding to develop more fluid use of the strategies. This is something I plan to incorporate into my teaching of this group in the coming academic year.

8.2 FUTURE IMPLICATIONS

Several findings from this research project merit further exploration within my practice. Firstly, I feel there is enough evidence from the students to support continued use of think-aloud modelling, but only when new material is presented. I plan to move the scaffolding process along and ask the students to think-aloud to each other more. This will allow me to provide feedback to enable further progress (Rosenshine & Meister, 1992). It will be interesting to see if they develop more expert-like thinking and a more fluid use of strategies with more time and opportunity for practice.

Secondly, with the lack of a discernable increase in attainment, I am interested in tracking the attainment of the cohort over the course of Year 13 to see if there is a delayed impact from the intervention. Given the aim of this research was related to low attainment when responding to higher order questions in my students, I plan to look closely at cohort attainment in AO2 and AO3 questions in their external examinations. While there will be too many variables to draw any

concrete conclusions from a potential increased attainment in relation to the intervention, I might be able to modify my practice further with that evidence.

Thirdly, the findings also highlighted the importance of promoting both metacognition and self-efficacy in the classroom. Further exploration of both, in relation to my practice and in the wider school context, feels necessary. Plans are already underway to incorporate more teacher modelling into departmental practice following dissemination of my initial findings and the advocacy of my colleague in a departmental meeting. More needs to be done to deepen practitioner understanding of the effect of metacognition and self-efficacy on science achievement, to be used in conjunction with the modelling practice.

While the small-scale nature of this project suggests that only 'fuzzy generalizations' can be made from these findings (Bassey, 1999), the intervention used is highly adaptable to many different settings. The context of my students is unique, but they are not unique in their struggle to use HOTS to answer higher order questions. The comments from my colleague in relation to the student experience of repeated failure without improvement in more challenging questions could be applied to any number of students in a wide variety of contexts. The hope is that the benefits of expert teacher modelling, using explicit strategies to approach problems and demonstrating their own fallibility, will invite professional discourse as to how the findings can be applied in another context (Bassey, 1999).

9 REFERENCES

- Allen, K. D., & Hancock, T. E. (2008). Reading comprehension improvement with individualized cognitive profiles and metacognition. *Literacy Research and Instruction, 47*(2), 124–139.
<https://doi.org/10.1080/19388070801938320>
- Alonzo, J., Basaraba, D., Tindal, G., & Carriveau, R. S. (2009). They read, but how well do they understand?: An empirical look at the nuances of measuring reading comprehension. *Assessment for Effective Intervention, 35*(1), 34–44.
<https://doi.org/10.1177/1534508408330082>
- Agarwal, P. K. (2019). Retrieval practice & Bloom’s taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology 111*(2), 189-209.
<https://doi:10.1037/edu0000282>
- Anderson, R. C., & Pearson, P. D. (1988). A schema-theoretic view of basic processes in reading comprehension. In P. L. Carrell, J. Devine & D. E. Eskey (Eds.), *Interactive Approaches to Second Language Reading* (pp. 37–55). Cambridge University Press.
<https://doi.org/10.1017/CBO9781139524513.007>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom’s taxonomy of educational objectives* (Abridged ed.). Longman.
- AQA. (2014, Nov). *Command words*. <https://www.aqa.org.uk/resources/science/as-and-a-level/biology-7401-7402/teach/command-words>
- AQA. (2021) *AS and A-Level biology* (Version 1.5). AQA
<https://www.aqa.org.uk/subjects/science/as-and-a-level/biology-7401-7402/specification-at-a-glance>
- AQA. (2007). *GCE AS and A Level biology* (Version 1.4). AQA

- Aubrey, K., & Riley, A. (2022). *Understanding and using educational theories* (3rd edition.). SAGE Publications.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioural change. *Psychological Review*, 84(2), 191-215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117–148. https://doi.org/10.1207/s15326985ep2802_3
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. W.H. Freeman.
- Bartlett, F. C. S. (1958). *Thinking: an experimental and social study*. Basic Books.
- Bassey, M. (1999). *Case study research in educational settings*. OUP McGraw-Hill Education.
- Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction* (2nd ed.). The Guilford Press.
- Ben-Chaim, D., Ron, S., & Zoller, U. (2000). The disposition of eleventh-grade science students toward critical thinking. *Journal of Science Education and Technology*, 9(2), 149–159. <https://doi.org/10.1023/A:1009474023364>
- British Educational Research Association. (2024). *Ethical guidelines for educational research* (5th ed.). www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2024
- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1, Cognitive domain*. Longmans.
- Bjork, E. L., & Bjork, R. A. (2014). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher and J. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (2nd ed.). (pp. 59-68). Worth Publishing

- Bohn, C. M., Roehrig, A. D., & Pressley, M. (2004). The first days of school in the classrooms of two more effective and four less effective primary-grades teachers. *The Elementary School Journal*, 104(4), 269–287. <https://doi.org/10.1086/499753>
- Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn: brain, mind, experience, and school* (Expanded ed.). National Academy Press.
- Brooks, R., te Riele, K., & Maguire, M. (2014). *Ethics and education research* (First Edition.). SAGE.
- Brown, P. L., & Concannon, J. P. (2016). Students' perceptions of vocabulary knowledge and learning in a middle school science classroom. *International Journal of Science Education*, 38(3), 391-408. <https://doi.org/10.1080/09500693.2016.1143571>
- Brown, B. A., Donovan, B., & Wild, A. (2019). Language and cognitive interference: How using complex scientific language limits cognitive performance. *Science*, 103(4), 750–769. <https://doi.org/10.1002/sci.21509>
- Cammies, C., Cunningham, J. A., & Pike, R. K. (2024). Not all Bloom and gloom: assessing constructive alignment, higher order cognitive skills, and their influence on students' perceived learning within the practical components of an undergraduate biology course. *Journal of Biological Education*, 58(3), 588–608. <https://doi.org/10.1080/00219266.2022.2092191>
- Clark, T., Foster, L., Sloan, L., Bryman, A., & Bryman, A. (2021). *Bryman's social research methods* (Sixth edition.). Oxford University Press.
- Cohen, L., Manion, L., & Morrison, K. R. B. (2018). *Research methods in education* (Eighth edition). Routledge.

- Cordingley, P. (2015). The contribution of research to teachers' professional learning and development. *Oxford Review of Education*, 41(2), 234–252.
<https://doi.org/10.1080/03054985.2015.1020105>
- Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE Life Sciences Education*, 7(4), 368–381. <https://doi.org/10.1187/cbe.08-05-0024>
- Cuban, L. (1984). Policy and research dilemmas in the teaching of reasoning: Unplanned designs. *Review of Educational Research*, 54(4), 655–681.
<https://doi.org/10.3102/00346543054004655>
- Central University Research Ethics Committee. (2022). CUREC Best practice guidance 04_Version 3.0: Research involving competent youths.
<https://researchsupport.admin.ox.ac.uk/files/bpg04competentyouthspdf-0>
- Dabarera, C., Renandya, W. A., & Zhang, L. J. (2014). The impact of metacognitive scaffolding and monitoring on reading comprehension. *System (Linköping)*, 42(1), 462–473.
<https://doi.org/10.1016/j.system.2013.12.020>
- Das, S., Das Mandal, S. K., & Basu, A. (2022). Classification of action verbs of Bloom's taxonomy cognitive domain: An empirical study. *Journal of Education*, 202(4), 554–566.
<https://doi.org/10.1177/00220574211002199>
- Denscombe, M. (2021). *The good research guide: research methods for small-scale social research* (Seventh edition.). Open University Press.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: verbal reports as data* (Rev. ed.). The MIT Press

- Facione, P. A., & Facione, N. C. (2013). Critical thinking for life: Valuing, measuring, and training critical thinking in all its forms. *Inquiry*, 28(1), 5–25.
<https://doi.org/10.5840/inquiryct20132812>
- Fang, Z. (2006). The language demands of science reading in middle school. *International Journal of Science Education*, 28(5), 491–520. <https://doi.org/10.1080/09500690500339092>
- Farrimond, H. (2017). The Ethics of Research. In D. Wyse, N. Selwyn, E. Smith, & L. Suter (Eds.) *The BERA/SAGE Handbook of Educational Research* (pp. 72-89). London: Sage Publications.
- Fisher, P. J., & Blachowicz, C. L. (2013). A few words about math and science. *Educational Leadership*, 71(3), 46-51.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. *The Nature of Intelligence*, 12, 231–235.
- Furst, E. J. (1981). Bloom’s taxonomy of educational objectives for the cognitive domain: Philosophical and educational Issues. *Review of Educational Research*, 51(4), 441–453.
<https://doi.org/10.3102/00346543051004441>
- Gagné, R. M. (1965). *The conditions of learning*. New York: Holt, Rinehart and Winston.
- Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. *Journal for Research in Mathematics Education*, 23(3), 242–273.
<https://doi.org/10.2307/749120>
- Holman, J. & Yeomans, E. (2018). Improving secondary science: Guidance report. London: Education Endowment Foundation. [Secondary-Science-v2.96-WEB.pdf](https://d2tic4wvo1iusb.cloudfront.net/Secondary-Science-v2.96-WEB.pdf)
[\(d2tic4wvo1iusb.cloudfront.net\)](https://d2tic4wvo1iusb.cloudfront.net/)
- Hwang, G.-J., Lai, C.-L., Liang, J. C., Chu, H.-C., & Tsai, C.-C. (2018). A long-term experiment to investigate the relationships between high school students’ perceptions of mobile learning

- and peer interaction and higher-order thinking tendencies. *Educational Technology Research and Development*, 66(1), 75–93. <https://doi.org/10.1007/s11423-017-9540-3>
- Irvine, J. (2017). A Comparison of revised Bloom and Marzano’s new taxonomy of learning. *Research in Higher Education Journal*, 33.
- Jackson, V. (2016). Applying the think-aloud strategy to improve reading comprehension of science content. *Current Issues in Education*, 19(2), 1–35. [Applying the Think-Aloud Strategy to Improve Reading Comprehension of Science Content | Current Issues in Education \(asu.edu\)](https://doi.org/10.1007/s11423-017-9540-3)
- Jarvis, P. (1999). *The practitioner-researcher: developing theory from practice*. Jossey-Bass.
- Jones, K. O., Harland, J., Reid, J. M. V., & Bartlett, R. (2009). Relationship between examination questions and Bloom’s taxonomy. *2009 39th IEEE Frontiers in Education Conference*, 1–6. <https://doi.org/10.1109/FIE.2009.5350598>
- Jordan, R. R. (1997). *English for academic purposes: a guide and resource book for teachers*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511733062.017>
- Karpicke, J. D., & Roediger, H. L. I. (2008). Critical importance of retrieval for learning. *Science*, 319(5865), 966–968. <https://doi.org/10.1126/science.1152408>
- King, F. J., Goodson, L., & Rohani, F. (1998). Higher order thinking skills: Definition, teaching strategies, assessment. *Center for the Advancement of Learning and Assessment Florida State University*. [higher-order-thinking-skills .pdf \(wordpress.com\)](https://www.claa.fsu.edu/wp-content/uploads/2017/06/higher-order-thinking-skills.pdf)
- Lee, J., & Choi, H. (2017). What affects learner’s higher-order thinking in technology-enhanced learning environments? The effects of learner factors. *Computers and Education*, 115, 143–152. <https://doi.org/10.1016/j.compedu.2017.06.015>
- Krathwohl, D. R. (2002) A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-218, https://doi.org/10.1207/s15430421tip4104_2

- Kreitzer, A., & Madaus, G. (1994). Empirical investigations of the hierarchical structure of the taxonomy. In L. Anderson & L. Sosniak (Eds.) *Bloom's taxonomy: A forty-year retrospective* (pp. 64-81). University of Chicago Press.
- Leighton, J. P. (2017). *Using think-aloud interviews and cognitive labs in educational research*. Oxford University Press.
- Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into Practice*, 32(3), 131-137. <https://doi.org/10.1080/00405849309543588>
- Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading & Writing Quarterly*, 19(2), 119–137. <https://doi.org/10.1080/10573560308223>
- Liu, J., Ma, Y., Sun, X., Zhu, Z., & Xu, Y. (2022). A systematic review of higher-order thinking by visualizing its structure through HistCite and CiteSpace software. *The Asia-Pacific Education Researcher*, 31(6), 635–645. <https://doi.org/10.1007/s40299-021-00614-5>
- Long, R. (2017). *GCSE, AS and A level reform (England)*. Briefing Paper Number 06962, 31 March 2017. <https://researchbriefings.files.parliament.uk/documents/SN06962/SN06962.pdf>
- Magno, C. (2010). The role of metacognitive skills in developing critical thinking. *Metacognition and Learning*, 5(2), 137–156. <https://doi.org/10.1007/s11409-010-9054-4>
- Maier, N. R. F. (1933). An Aspect of Human Reasoning. *British Journal of Psychology. General Section*, 24(2), 144–155. <https://doi.org/10.1111/j.2044-8295.1933.tb00692.x>
- Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives* (2nd ed.). Corwin.
- Menter, I., Elliot, D., Hulme, M., Lewin, J., & Lowden, K. (2011). *A guide to practitioner research in education* (1st ed.). SAGE Publications.

- Mkimbili, S. T. (2022). Do biology syllabi provide opportunities for secondary school students to engage with critical thinking skills? *Journal of Biological Education*, 58(2), 444–459.
<https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1080/00219266.2022.2067582>
- Miri, B., Ben-Chaim, D., & Zoller, U. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in Science Education*, 37(4), 353–369.
<https://doi.org/10.1007/s11165-006-9029-2>
- Muijs, D. and Bokhove, C. (2020). *Metacognition and self-regulation: Evidence review*. Education Endowment Foundation. [Metacognition and self-regulation review.pdf](https://www.educationendowmentfoundation.org.uk/evidence-reviews/metacognition-and-self-regulation-review)
 [\(d2tic4wvo1iusb.cloudfront.net\)](https://d2tic4wvo1iusb.cloudfront.net)
- Muijs, D., & Reynolds, D. (2018). *Effective teaching: evidence and practice* (4th edition.). SAGE.
- Newmann, F. M. (1990). Higher order thinking in teaching social studies: A rationale for the assessment of classroom thoughtfulness. *Journal of Curriculum Studies*, 22(1), 41–56.
<https://doi.org/10.1080/0022027900220103>
- Nunes, T., Bryant, P. and Barros, R. (2012) The development of word recognition and its significance for comprehension and fluency. *Journal of Educational Psychology*, 104(4), 959–973, <https://doi:10.1037/a0027412>
- Nunes, T., Bryant, P., Strand, S., Hillier, J., Barros, R., Miller-Friedmann, J. (2017). *Review of SES and science learning in formal educational settings: A report prepared for the EEF and the Royal Society*. Education Endowment Foundation.
[Review of SES and Science Learning in Formal Educational Settings.pdf](https://www.educationendowmentfoundation.org.uk/evidence-reviews/review-of-ses-and-science-learning-in-formal-educational-settings)
 [\(d2tic4wvo1iusb.cloudfront.net\)](https://d2tic4wvo1iusb.cloudfront.net)
- OECD (2017). PISA 2015 Science Framework. In *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*.
<https://doi.org/10.1787/9789264281820-3-en>

- Pólya, G. (1957). *How to solve it: A new aspect of mathematical method* (2nd ed.). Princeton University Press.
- Quigley, A., Mujs, D., & Stringer, E. (2021). *Metacognition and self-regulated learning: Guidance report*. Education Endowment Foundation. [EEF Metacognition and self-regulated learning.pdf \(d2tic4wvo1iusb.cloudfront.net\)](https://www.eef.org.uk/media/1000/EEF_Metacognition_and_self-regulated_learning.pdf)
- Resnick, L. B. (1987). Education and learning to think. *National Academy Press*. [Education and Learning to Think | The National Academies Press](https://www.nationalacademies.org/pubs/education-and-learning-to-think/)
- Rosenshine, B., & Meister, C. (1992, Apr). The use of scaffolds for teaching higher-level cognitive strategies. *Educational Leadership*, 49(7), 26–33.
- Scheper-Hughes, N. (2000). Ire in Ireland. *Ethnography*, 1(1), 117–140. <https://doi.org/10.1177/14661380022230660>
- Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, 36(1), 111–139. <https://doi.org/10.1007/s11165-005-3917-8>
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460–475. <https://doi.org/10.1006/ceps.1994.1033>
- Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, 328(5977), 450–452. <https://doi.org/10.1126/science.1182597>
- Stanny, C. (2016). Reevaluating Bloom’s taxonomy: What measurable verbs can and cannot say about student learning. *Education Sciences*, 6(4), 37. <https://doi.org/10.3390/educsci6040037>
- Sun, H., Xie, Y., & Lavonen, J. (2022). Exploring the structure of students’ scientific higher order thinking in science education. *Thinking Skills and Creativity*, 43, 100999. <https://doi.org/10.1016/j.tsc.2022.100999>

- Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In *Cognitive Load Theory* (pp. 29–47). Cambridge University Press. <https://doi.org/10.1017/CBO9780511844744.004>
- Torrance Jenkins, R. (2017). Using educational neuroscience and psychology to teach science. Part 1: A case study review of cognitive load theory (CLT) and Cognitive Acceleration through Science Education (CASE). *School Science Review*, 99(367), 93-103.
- Tracy, S. J. (2010). Qualitative quality: Eight “big-tent” criteria for excellent qualitative research. *Qualitative Inquiry*, 16(10), 837–851.
<https://doi.org/10.1177/1077800410383121>
- UNEP. (2012). *21 Issues for the 21st Century: Result of the UNEP Foresight Process on Emerging Environmental Issues*. United Nations Environment Programme (UNEP).
www.unep.org/resources/report/21-issues-21st-century-results-unep-foresight-process-emerging-environmental
- Vacca, R. T., Vacca, J. A. L., & Mraz, M. E. (2014). *Content area reading: literacy and learning across the curriculum* (Eleventh ed.). Pearson.
- Vygotsky, L. S. (1978) *Mind in society: Development of higher psychological processes*. Harvard University Press. <https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.2307/j.ctvjf9vz4>
- Walsh, P., Owen, P., & Mustafa, N. (2021). The creation of a confidence scale: The confidence in managing challenging situations scale. *Journal of Research in Nursing*, 26(6), 483-496.
<https://doi.org/10.1177/1744987120979272>
- Willingham, D. T. (2021). *Why don't students like school?: A cognitive scientist answers questions about how the mind works and what it means for the classroom* (2nd ed.). Wiley.

- Wineburg, S., & Schneider, J. (2009). Was Bloom's taxonomy pointed in the wrong direction? *Phi Delta Kappan*, 91(4), 56–61. <https://doi.org/10.1177/003172171009100412>
- Yerdelen-Damar, S., & Pesman, H. (2013). Relations of gender and socioeconomic status to physics through metacognition and self-efficacy. *The Journal of Educational Research*, 106(4), 280–289. <https://doi.org/10.1080/00220671.2012.692729>
- Yin, R. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.
- Zheng, A. Y., Lawhorn, J. K., Lumley, T., & Freeman, S. (2008). Assessment: application of Bloom's taxonomy debunks the "MCAT myth." *Science*, 319(5862), 414-415. <https://doi.org/10.1126/science.1147852>
- Zoller, U. (1993). Are lecture and learning compatible? Maybe for LOCS: Unlikely for HOCS. *Journal of Chemical Education*, 70(3), 195–197. <https://doi.org/10.1021/ed070p195>

10 APPENDICES

10.1 APPENDIX A : FULL VERSION OF TABLE 5

Question type	Demand of question (adapted from AQA, 2014)	Link to revised taxonomy using action verbs (Das et al., 2022)	Command words included in this category
Making a judgement – using evidence	Either making an informed judgement, judging from available evidence or supporting a case with evidence.	Evaluate	Is the conclusion correct? Justify Assess why Consider whether the graph supports a conclusion What can you conclude? Use data to justify Evaluate conclusion Evaluate Use information to explain conclusion (put here because a judgement would be made about data to link to conclusion)
Suggest	Present a possible cause	Not in list	Suggest Suggest how Suggest and explain Suggest improvement Suggest, use calculation Use the data to suggest how
Explain	Give reasons	Comprehension	Explain Explain how Explain why Describe and explain Explain result Explain how and why
Dealing with data		Application K & C, but with data	Calculate Describe relationship (on graph) Describe the evidence (data)
Describe	Give an account of	Knowledge & Comprehension	Describe Describe how Describe what Give and describe
Recall		Knowledge, except for identify, which spans knowledge and comprehension	Draw Draw and identify Which What Give Name State (use table)

10.2 APPENDIX B: QUESTIONNAIRE

1. Rate your confidence level

	No Confidence	Little Confidence	Some confidence	Confidence	High Confidence
In this topic before seeing the questions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In this topic after answering the questions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Answering questions that require you to read a lot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Answering questions that require you to interpret data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Answering questions that contain unfamiliar words	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2. What went well for you in this set of questions?

Enter your answer

3. What did you find challenging about this set of questions, as a whole?

Enter your answer

4. What are the reasons you lost marks? Put the biggest reason first.

Enter your answer

5. How might you try to approach questions differently in the future?

Enter your answer

10.3 APPENDIX C: ADDITIONAL POST-INTERVENTION QUESTIONNAIRE QUESTIONS

6. Since January, how has your confidence in answering more challenging questions changed?

Enter your answer

7. What has had the biggest impact on your confidence in answering questions, since January?

Enter your answer

10.4 APPENDIX D: EXAMPLE OF CODING PROCESS USED IN ANALYSIS OF QUESTIONNAIRE

What are the reasons you lost marks? Put the biggest reason first.

Cell mem	Imm	Eco	April	Post
<p>-misreading the question, wording, using all the information</p> <p>-silly mistakes, lack of detail/ elaboration</p> <p>-Not understanding the question - for example no.1, or last question</p> <p>-lack of knowledge from missing previous questions and lack of revision</p> <p>-overall more revision needed, a lack of understanding of what the questions asking, keep focused</p>	<p>-vocabulary, understanding question, lack of detail</p> <p>-lack of knowledge- it would've been easier to apply if i fully knew all the types of immunity, using key vocab and not reading the question properly (after reading the mark scheme it seemed like i overcomplicated it) it</p> <p>- Didn't answer the last 2 questions, Didn't redo my calculation (i did it too complicated and lost marks)</p> <p>- lack of time to read question and understand it</p> <p>- Understanding, maths</p>	<p>not really knowing the content well enough, i couldn't completely differentiate between different subtopics, they all interlink and merge in my head</p> <p>- Not knowing the mark scheme caused me to lose a high number of marks, as well as describing investigative procedures/investigations</p>	<p>-Vocab, detail, understanding of question</p> <p>- lack of detail due to a lack for deeper understanding and continued revision in application questions, rushing questions near the end of the paper</p> <p>- Lack of detail when directly answering the question, Silly maths errors</p> <p>- lack of knowledge</p> <p>- silly mistakes, understanding what they are asking for and using the question to help guide me.</p>	<p>- didn't use specific vocab, not enough detail/expansion of thought, understanding of topic, lack of knowledge, graph skills</p> <p>- silly mistakes and lack of detail or expansion, also vocab and not understanding the questions</p> <p>- Lack of knowledge</p> <p>- overthinking the question</p> <p>- keeping focus and not losing track of the question, not fully understanding the questions</p>
<p>Definitely a pattern of not understanding the Qs here. Also – depth of knowledge – linked to detail as well.</p>	<p>Understanding Q</p> <p>Depth of knowledge (detail)/vocab</p> <p>Difficult calculation</p> <p>Timing</p>	<p>Mostly cited lack of knowledge (perhaps depth) here. Interesting comment about 'not knowing m.s.). Could link to practice or knowing m.s. for similar types of Qs. (only 2 students)</p>	<p>Lack of detail; understanding Q;</p> <p>Knowledge and applying knowledge.</p>	<p>Not fully understanding Q, though not always 1st reason given. Knowledge in terms of detail and vocab and their use.</p>

Summary:

Main things mentioned: Not understanding the Q (what the Q is asking for), Lack of knowledge/understanding of the content, lack of detail or use of specific vocabulary when answering the Q. By post-inter test – potentially mentioning understanding Q as a barrier lower down in the reasons given.

10.5 APPENDIX E: GROUP INTERVIEW QUESTIONS

1. Since participating in the project, you have been thinking about which types of questions are most challenging – has that changed at all since the start?
2. What impacts the difficulty of a question more (the topic, the wording, or the type of Q)?
3. What have you found useful about reviewing your tests and coding the questions afterwards, as we did yesterday?
4. What have you found less useful about that?
5. How are you finding the experience of your teachers modelling their thinking when answering a question?
6. What is useful about the strategies (SKLIS ME) you've been given?
7. What is less useful about the SKLIS ME?
8. Will you keep using the strategies? Why or why not?
9. How has participating in the study changed how you think about questions and how you respond to them?
10. How confident do you now feel about approaching harder questions, such as 'Suggest' 'Evaluate', or data interpretation, compared to at the start of the project?
11. Is there anything else you would like to say about the process you have been involved in?

10.6 APPENDIX F: EXAMPLE OF CODING USED FOR INTERVIEW QUESTIONS

12. Since participating in the project, you have been thinking about which types of questions are most challenging – has that changed at all since the start?

- E: 'Hard ones have gotten easier now because I know how to approach them, but they are definitely still the hard ones.' Hard ones = interpretation of data, suggest...actually the interpretation of data aren't too bad, but it's the wordy ones.
- C: I'm sort of the opposite. It's kind of flipped for me. So at the start of the term, I was felt like interpreting data was not that hard in comparison to something like evaluate or suggest. But now I think, interpreting data are just the hardest ones. And then the others are pretty easy to do when you just approach it in the right way.
- B: Interpreting data, it's harder than I thought it was going to be. I think it's a bit of a step up from GCSE and the recall questions are actually harder than I thought they were going to be as well. Yeah, yeah, like that, quite. Like what they're asking you is - you need to refine things down more than I thought you did, I think, at the start.
- A: Yeah, I mean, like, I think I still find, like, the interpreting data ones, hard. And, like, the wordy ones are also quite hard, but, like, they've gotten more, like, manageable, like, from all the... 'How to approach things?'
- D: I think the wordy ones are hard for me, it's like I overthink a question and I don't write about the question, [laughter] no, OK. I go off topic a bit.' When probed further about overthinking – puts her off because she doesn't know where to start? Yes.

10.7 APPENDIX G: QUESTIONS FROM TEACHER INTERVIEWS

Pre-intervention

1. How confident do you think A-Level students are in answering harder questions?
2. Could you categorise the types of questions students struggle with the most?
3. How do you deal with these types of difficulties?

Post-intervention

1. Share data. Do these data align with your thoughts/opinions of what students find challenging about answering (harder) questions?
2. Have you used the post-assessment coding for students to assign a reason to their lack of marks before? What are the positives/negatives of that process? How did you find it more recently?
3. How are you finding the intervention practice?
4. What effect has this practice had on:
 - a. Your teaching?
 - b. Your students' success in answering questions?
 - c. Your students' learning?
5. Is this something you would continue to do in lessons? Why or why not?
6. How would you suggest this practice be used in the future? Any changes?
7. Is there anything else you would like to share about your involvement in this process?

10.8 APPENDIX H: EXAMPLE OF ANNOTATED QUESTION USED TO MODEL THINK-ALOUD STRATEGIES DURING THE INTERVENTION

Q1. I do

The fruit fly is a species of small insect.

The fruit fly has a gene that codes for an enzyme called alcohol dehydrogenase (AD). AD catalyses the breakdown of alcohol when alcohol is in the insects' food.

The gene coding for AD has two alleles, AD^F and AD^S.

- (a) The enzyme encoded by the AD^F allele catalyses the breakdown of alcohol faster than the enzyme encoded by the AD^S allele. Suggest why.

- alleles code for slightly different enzymes
↳ different sequence of DNA bases → leads to diff. 1^o structure
- 2 enzymes have different 3^o structure
- may mean that AD^F binds with the alcohol more readily due to active site being better complement/better fit
- so catalysis happens faster

2
(3)