UNIVERSITY OF CAMBRIDGE
Faculty of Education

*ISSN 2043-8338*

*Journal of Trainee Teacher Education Research*

# A critical analysis of how the use of diagnostic questions as formative assessment probes can enhance teaching and learning in Year 8 lessons on infectious diseases

**Katherine Webb**

(PGCE Secondary Science (Biology), 2019-2020)

email: katherine.s.webb96@gmail.com

## Abstract

*Formative assessment probes are a well-established pedagogical tool, used to uncover student understanding via low-stakes, diagnostic questioning, which informs feedback. Despite the development and use of question banks of formative assessment probes, the field lacks research supporting the outcomes of the tool. This paper investigates effective use of formative assessment probes and student perceptions of the technique. Results following the use of formative assessment probes throughout a Year 8 scheme of work suggested formative assessment probes can enhance teaching and learning. However, probes must be carefully devised in order to effectively evidence learning. Qualitative data suggested that consistent use of the technique resulted in positive student perceptions and potentially increased motivation to learn, as students perceived probes as useful. Probes elicited evidence of scientific explanations as well as factual recall to provide powerful formative assessment.*

# A critical analysis of how the use of diagnostic questions as formative assessment probes can enhance teaching and learning in Year 8 lessons on infectious diseases

**Katherine Webb**

## Introduction

'Formative assessment' is a term familiar to secondary school teachers. With strong evidence that formative assessment enhances teaching and learning (Black & Wiliam, 1998), various methods have been developed to facilitate the process (Wiliam, 2011a). Formative assessment entails a continuous, responsive process of assessment activities to inform teaching and provide students with meaningful feedback (Trauth-Nare & Buck, 2011). In science education, formative assessment probes can be used prior to and during instruction to uncover student thinking, evidencing misconceptions to guide feedback (Dokter, Pompea, Sparks & Walker, 2010; Keeley, 2018).

Formative assessment probes consist of low-stakes, diagnostic questioning, which guides feedback and teaching (Dokter et al., 2010). Probes often consist of a multiple-choice question (MCQ) tackling factual recall, followed by a second question concerned with an explanation of the scientific concept (Bulunuz, Bulunuz, Karagoz & Tavsanli, 2016). This provides a two-tier approach, with both tiers equally important in evidencing learning. Probes often utilise MCQ components to ensure rapid completion, and evidence is accessible to teachers in real time. As a reflection of the efficiency MCQs offer, teachers are increasingly using them for assessment (Gierl, Bulut, Guo & Zhang, 2017), but research regarding success of MCQs specifically in terms of formative assessment probes that integrate this format alongside a second tier of explanation appears to be lacking. The second tier adds additional insight into student learning, but the accuracy and usefulness of this component again appears to lack research. Similarly, the perceptions of UK secondary school students regarding this technique are unknown.

This study draws upon academic literature and classroom-based research to investigate the use of formative assessment probes in science classrooms. Formative assessment probes identify common misconceptions, providing information to be used formatively (Keeley, 2015). Scientific misconceptions have afforded much research due to their importance in the construction of scientific narratives (Driver, Squires, Rushworth & Wood-Robinson, 2015) and the difficulty of rectifying these naïve understandings of the world (Queloz, Klymkowsky, Stern, Hafen & Köhler, 2017). Therefore, quick, simple, pre-prepared probes are certainly appealing. This study investigates the impact of probes upon teaching and learning.

A group of year 8 students of low to middle attainment (n = 32) were subject to an intervention of research-supported formative assessment probes as part of an action research investigation. The topic of 'Infectious Diseases' provided a subject for which common misconceptions had been identified by research to aid construction of probes. Summative assessment was used to evidence the potential of these probes to enhance learning, alongside evidence regarding student perceptions provided by questionnaire responses and supported by classroom observations. To analyse the effects of probes in enhancing teaching practice, student responses to formative assessment probes were reviewed to confirm their accuracy in evidencing student learning.

The importance of this study is exemplified by ongoing recognition of the difficulties presented by delivering effective formative assessment (Andersson & Palm, 2018). Formative assessment has been identified as an important tool for enhancing teaching and learning (Black & Wiliam, 1998). Further, in science education specifically, research suggests identification of pre-existing learning and established misconceptions is important to providing quality teaching (Driver et al., 2015). Providing teachers with quality tools that enable this and work in real classrooms makes progress towards addressing what Black and Wiliam (1998) call a 'poverty of practice' regarding formative assessment.

For the purposes of this paper in addressing overarching themes of enhancing teaching and learning, 'learning' is defined as the active process of acquiring new knowledge or behaviours of relative permanence. Therefore, factors that enhance learning are those that assist students in acquiring knowledge and developing behaviours that support this. 'Teaching' is defined as the process of supporting student progression towards specified learning objectives in a way that is considerate of and reactive to student needs, as defined by Hirst (1975). Therefore, factors that enhance teaching contribute not only to empirical evidence of learning, but also to an understanding of student needs to assist planning and facilitate learning.

## Literature Review

This literature review outlines existing research regarding formative assessment in secondary science education and the use of formative assessment probes for this purpose. Literature is critically analysed to inform the planning process for this study.

### Formative assessment

In order to understand the potential of formative assessment probes to enhance teaching and learning, the foundations of formative assessment are first outlined to establish the role of probes in achieving perceived benefits. As the result of an extensive literature review to clarify the importance of formative assessment, Black and Wiliam (1998) argue formative assessment in science education must be 'integral' to teaching, and that inclusion of effective formative assessment enhances teaching and learning. Black and Wiliam (1998) identify five strategies via which effective formative assessment can be achieved, re-worded / paraphrased here as:

1. Engineering effective classroom discussions, activities, and learning tasks to elicit evidence of learning
2. Providing feedback that moves learning forward
3. Clarifying, sharing, and understanding learning intentions/criteria for success
4. Activating learners as owners of their learning
5. Activating learners as instructional resources for one another

As outlined by Wiliam (2018), activities via which these strategies can be achieved comprise a range of tasks, including self-assessment, group work, or scaffolded worksheets to name but a few in addition to formative assessment probes. These activities have the potential to enhance teaching by offering useful formative assessment. Not all activities fall under the realm of every strategy, but formative assessment probes have been shown to elicit evidence of learning (Sadler & Sonnert, 2016), highlighting misconceptions to provide feedback that moves learning forward (Bulunuz & Bulunuz, 2013).

Despite long-standing consensus in the field of education that formative assessment enhances teaching, exploring tools that enable *effective* formative assessment remains of great importance. In 1998, Black and Wiliam recognised a 'poverty of practice' with regard to formative methods – a problem still reflected in today's literature, as teachers encounter difficulties in devising effective techniques (Andersson & Palm, 2018). Although formative assessment has the potential to enhance teaching, realising this potential is a modern educational issue.

*Eliciting evidence of learning to achieve effective formative assessment*

The primary function of formative assessment probes is to elicit evidence of learning through diagnostic testing, often in the form of MCQs (Galvin, Simmie & O'Grady, 2015; Wiliam, 2011a, p.97). Scientific misconceptions – conceptual ideas that differ from the consensus view of the scientific community (Galvin et al., 2015) – are often rooted in sensory or social experiences, and are difficult to overcome (Driver et al., 2015; Queloz et al., 2017; Slaughter & Ting, 2010; Wind & Gale, 2015). Formative assessment probes are designed to provide teachers with feedback with regard to student misconceptions (a diagnostic purpose), and provide students with feedback via clarification of knowledge, to support the identification of steps towards progress (a formative purpose) (Van der Kleij, Vermeulen, Schildkamp, & Eggen, 2015). Table 1 presents a summary of how diagnostic assessment can be used effectively in formative assessment, based on Keeley (2018). As shown in Table 1, diagnostic testing becomes formative when used to guide feedback and teaching (Keeley, 2018). Consequently, this two-step process is reliant upon *accurate* evidencing of learning. The question, therefore, is whether the evidence provided by formative assessment probes is accurate enough to provide useful feedback.

| Classroom assessment | Purpose | Stage in lesson/sequence |
|---|---|---|
| Diagnostic | Identification of preconceptions or difficulties | Pre-instruction<br>During instruction |
| Formative | To provide information to give feedback and inform teaching | Pre-instruction<br>During instruction<br>Post-instruction |
| Summative | To evaluate student understanding | Post-instruction |

**Table 1: A summary of classroom assessment. Based on Keeley, 2018**

Research has established common misconceptions across a range of content (Driver et al., 2015), which can be used to produce diagnostic questions as the basis of formative assessment probes that elicit evidence of prior or developing understanding (Keeley, 2015; Sadler & Sonnert, 2016). With research supporting diagnostic testing to evidence misconceptions (Fuchs & Arsenault, 2017; Queloz et al., 2017; Wind & Gale, 2015), banks of formative assessment probes have been developed. For example, in 2005, Page Keeley first started writing about formative assessment probes (Keeley, Eberle & Farrin, 2005). As a pioneer of the tool, Keeley presented a design consisting of an MCQ, the question stem of which is contextualised within a 'real-life' situation, followed by space for written exploration of the concept, or a second MCQ to explain the first. In recognition of the strength of this two-tier design, many resources, including those from the University of York's Best Evidence Science Teaching (BEST) initiative, mirror this structure (University of York Science Education Group (UoYSEG), 2020). An example of such a design, created with this research project in mind, is shown below in Figure 1, developed following common misconceptions highlighted by Johnson and Bungum, (2013). Misconceptions are annotated, and the expected response is shown in green.

Catching a cold

Question is put into context

There is a cold going around the school and lots of students are feeling unwell. The headteacher is looking for advice for how to stop it spreading. Four students suggest different ideas. Who do you most strongly agree with?

**Alice** – We should wear extra jumpers to stop us getting the cold. *The cold virus is not reliant upon cool temperatures, nor is it caused by them*

**Peter** – Everyone should take antibiotics, just in case. *The common cold is caused by a virus, which cannot be treated with antibiotics*

**Mohammad** – We should cover our mouths and noses when coughing and sneezing to stop the cold spreading.

**Kayla** – We should all huddle together for warmth. *Echoes Alice's answer, and indicates a lack of understanding regarding virus transmission*

Why did you choose that answer? *Opportunity for scientific explanation (Tier 2)*

…………………………………………………………………………………………………………

…………………………………………………………………………………………………………

**Figure 1: An example formative assessment probe related to the topic of infectious disease.**

However, although commonly used in formative assessment probes, the ability of MCQs to evidence learning is disputed. With guesswork a probable factor and known difficulties regarding construction of good MCQs (Brown & Abdulnabi, 2017; Gierl et al., 2017; Shin, Guo, & Gierl, 2019), much research advises that MCQs are not effective in achieving this goal (Roberts, 2006). For example, Bulunuz et al. (2016) identify problems in the fragmentary and discrete nature of the surface-level understanding MCQs evidence, supported by Paxton (2000), who emphasises the lack of application necessitated by MCQs, detrimental to the development of conceptual understanding. Arguably, learning, as a multi-faceted and dynamic process, contradicts the notion of condensed, discrete knowledge available among a selection of four responses.

Conversely, opposing research recognises that these limitations can be overcome by constructing questions that encourage problem solving and explanation (Brown & Abdulnabi, 2017; Gierl et al., 2017). Related to this, and affording much research, is the design of MCQ distractors. Distractors – incorrect MCQ responses, so called to reflect their purpose in 'distracting' students from the correct answer – that are not plausible can lead to selection of correct answers due to distractors being too obviously incorrect, rather than via scientific reasoning (Shin, Guo & Gierl, 2019), or can introduce new misinformation (Gierl et al*.,* 2017). Consequently, distractor options used in probes are commonly drawn from research regarding common misconceptions (UoYSEG, 2020). This approach based upon 'common errors' produces effective distractors that identify misconceptions (Shin, Guo & Gierl, 2019).

Furthermore, two-tiered probes provide either an additional MCQ focused upon scientific explanation, or space for written explanations of tier one selections (Keeley, 2013b). Student explanations may highlight misconceptions and areas for development, extending the evidence provided by probes beyond factual recall (Keeley, 2013b). This reflects the problem-solving and explanation skills required for the conceptual learning these probes attempt to evidence (Bulunuz et al., 2016).

According to the above research, well-constructed formative assessment probes should be successful in evidencing learning, and perceived limitations of the technique appear to be mitigated by careful planning and use. Based upon this research, a two-tiered approach as presented by Keeley and BEST will be used as interventions in this study, with recommendations regarding distractor design and opportunities to evidence student explanations taken into account.

In doing so, this research attempts to address an apparent gap in the literature surrounding formative assessment probes - their success is not widely studied in practice. Indeed, even large efforts to collate banks of probes lack evidence of success from beyond their founding institutions. The BEST initiative (University of York) uses extensive research to produce diagnostic questions for formative assessment (UoYSEG, 2020), yet the only research supporting the outcomes of BEST is produced by researchers affiliated with the project (Millar, 2016; Millar & Hames, 2003; Whitehouse, 2013; Whitehouse, 2014). This may entail bias, as the university would be ill-advised to publish negative outcomes. Similarly, Keeley's publications have been popular, but their success in practice is largely supported by research produced by Keeley herself (Keeley 2013a; 2013b; 2015; 2016; 2018), with only a small number of other studies supporting these findings (for example, Bulunuz, 2016). In terms of wider, peer-reviewed literature, these banks lack confirmation of success in eliciting evidence of learning that Black and William (1998) specify is key to formative assessment.

**Diagnostic questioning as a basis for feedback: opposing research and proven guidelines**

As established, to fulfil the objectives of formative assessment, teachers utilise diagnostic testing provided by probes to facilitate feedback that enables correction of identified misconceptions, serving a formative purpose (Keeley, 2018). Therefore, the type of feedback formative assessment probes provide should be considered and evaluated.

In research, use of formative assessment probes appears to mainly generate verbal feedback due to the diversity of responses probes produce. Galvin et al. (2015) conclude that feedback with regard to misconceptions follows a pedagogical cycle of recognition, reduction and removal, facilitated by feedback methods that prompt deep thinking and reasoning, such as argumentation. Argumentation is a verbal tool in which students use prior and developing understanding of concepts to collaboratively align their beliefs to that of scientific consensus (Duschl & Osborne, 2002). Importantly, argumentation can be used to clarify correct answers, which has been identified as a key feature of effective feedback (Attali, Latitusis & Stone, 2016). Research supports the use of argumentation for feedback with regard to formative assessment probes, finding that argumentation is effective in tackling misconceptions, as students reason their answers via discussion (Bulunuz & Bulunuz, 2013).

It should be noted that Bulunuz and Bulunuz's (2013) research was conducted on a small sample of pre-service teachers rather than secondary school students. Therefore, findings may not be reflective of secondary school classroom practice. However, argumentation was also found to be an effective form of feedback to move learning forwards in a group of fifth grade students (Keeley, 2013a), with further research-based recommendations to stimulate discussion following probes (Keeley, 2016). In terms of theoretical support, Vygotskian theories emphasise social interactions are essential to developing understanding within the zone of proximal development in which students gain the skills necessary to harness higher level cognition through collaboration (Vygotsky, 1978). Further research is now necessary to appreciate the impact of these methods in UK secondary school classrooms.

Overall, this research suggests that formative assessment probes promote discursive feedback that research and theories recognise as effective. In light of this research, formative assessment probe interventions of this study will be followed by verbal feedback through argumentation and scaffolded discussion as outlined by Keeley (2013a; 2016).

**Positive student perceptions: usefulness as a key to formative assessment success**

Thus far, existing research has evidenced ways in which formative assessment probes can be used to elicit evidence of learning and provide feedback that moves learning forward – key features of effective formative assessment identified by Black and Wiliam (1998). These are aspects the teacher can control, via careful construction of questions and feedback. However, the learning process lies with the student (Wiliam, 2011b) and positive effects of formative assessment upon learning are reliant upon student willingness to utilise feedback (Kyaruzi, Strijbos, Ufer & Brown, 2019). Student motivation to do so becomes significant.

Research suggests effective formative assessment can foster a growth mindset in which students believe they can use feedback to incrementally progress academically via self-improvement, regardless of current ability (Wiliam, 2011b). Behaviours such as feedback-seeking are proven traits of the growth mindset (Cutumisu, 2019) and are aligned with intrinsic motivation styles, in which students pursue academic gain to support their own beliefs, rather than being extrinsically motivated by factors external to their locus of control, such as grades or competition (Ryan & Deci, 2000).

A recent study of English and Maths students at secondary school evidenced links between intrinsic motivation and positive perceptions of formative assessment feedback: when perspectives of the formative assessment process were positive, intrinsic motivation was heightened, and learning was enhanced (evidenced via improved academic performance) (Van der Kleij, 2019). Van der Kleij (2019) identifies perceived usefulness of formative assessment as a key influencer upon positive student perceptions of formative assessment. This is supported by research from Harks, Rakoczy, Hattie, Besser and Klieme (2014), who conclude that when students perceive feedback as useful, feedback can positively influence progress. Conversely, a lack of recognition of the usefulness of the formative assessment process would be detrimental to student perceptions of the technique, with anticipated effects upon motivation and academic performance.

Therefore, student perceptions of formative assessment probes are important. Negative perceptions may reduce the technique's potential to enhance learning despite the supporting research discussed. This may be especially true of MCQ-based formative assessment probes as over-use of MCQ-style diagnostic testing for summative purposes and comparatively low formative assessment use (Roberts, 2006) may result in students failing to identify formative usefulness of probes. Summative testing is not designed to give rise to feedback (Keeley, 2018), but formative assessment probes use similar questioning styles to do exactly that, which may not be intuitive to students. Currently, research does not comment upon student perceptions of MCQ-based formative assessment probes, posing questions regarding whether students identify the usefulness of them, as this may impact their potential to enhance learning.

## Research outline

Existing research suggests that when designed in accordance with research guidelines and followed by scaffolded argumentation, probes can provide formative assessment to enhance teaching and learning. However, research from UK secondary school science classrooms to support this appears to be lacking. In addition, whilst formative assessment has the potential to enhance learning, this is dependent upon student perceptions of feedback, particularly in regard to perceived usefulness. Therefore, this study focuses upon the following research questions (Table 2):

| | |
|---|---|
| Research question 1 (RQ1) | How effective are formative assessment probes in achieving the goals of effective formative assessment? |
| Research question 2 (RQ2) | How are student's perceptions of formative assessment probes affected by their consistent use throughout a scheme of work? |

**Table 2: Main research questions**

Action research enables practitioner-driven enquiry to inform and evaluate teaching practice (McNiff, 2017, p.9). Specifically, the method of this study follows notional action research: a concern is identified; a study to address this is devised; and the data gathered is evaluated with a view to

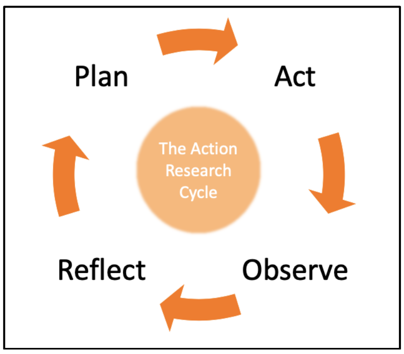


**Figure 2: An action research cycle. Based on McNiff, 2017.**

modification in further cycles (McNiff, 2017, p.11). Due to time constraints, only the first cycle of the action research process (depicted above in Figure 2) was completed.

## Methods and Methodology

Research was conducted during the second professional placement of the Post-Graduate Certificate of Education (PGCE) provided by the University of Cambridge. Seven of the eight lessons within this scheme of work were completed in regular classroom conditions, whilst lesson eight and the final assessment were completed online due to school closures.

### Research Group

The study was conducted at a mixed gender secondary school located in Cambridgeshire, England. The class in question are a low to middle attaining group of 32 students (n=32; 15 male, 17 female) following a spiral curriculum developed by the school for Key Stage 3 study. Full class numbers were infrequently achieved due to absence and school closures, entailing sample sizes of 20 to 31 students.

Of particular interest to this study is the subject content. The topic of ‘Infectious Diseases’ is not currently included in the National Curriculum guidelines but has been included at this level in recognition of its importance at a higher academic level and socially. Studies reveal that common misconceptions regarding disease persist at undergraduate level (Yekeen et al., 2017). Combined with research identifying links between understanding of disease and personal health (Hanson & Gluckman, 2011; Kilgour, Matthews, Christian & Shire, 2015; Pelikan, Ganahl & Roethlin, 2018), the importance of using techniques that address misconceptions for this topic is clear.

### Attainment Data

Following a previous topic, students completed a standardised summative assessment, with foundation and higher tiers assigned to represent prior attainment. A comparable summative assessment was completed at the end of the intervention lessons. Tests consisted of questions collated from Exampro (Exampro, 2020), or written by the experienced head of science, and had been piloted with other classes to confirm they produced attainment data reflective of student ability. The outcomes of the prior-intervention assessment were used as baseline data of attainment *without* formative assessment probes, and the final assessment was used as an indication of the effects of probes upon learning.

## Classroom observations

An intervention was observed by two other teachers (one pre-service, one in-service). Each teacher observed three students (n=6), with a range of abilities, behaviour and special educational needs represented across the sample. Observations were guided by the checklist protocol developed by Skinner, Kindermann and Furrer (2009). This checklist presents observers with statements related to themes of behavioural and emotional engagement or disaffection. If these identifiers were demonstrated by the student in question, the observer selected this statement, adding further descriptive narrative if necessary. This ensured that qualitative data was useful in contributing to conclusions regarding motivation and engagement. Using a research journal of my own reflections upon lessons, I also made observations regarding the influence of the tool upon personal practice and planning.

## Pre- and post-intervention questionnaires

A questionnaire was designed to analyse student perceptions of the formative assessment process using probes. Using a Likert scale, students indicated to what extent they agree with statements. Low scoring responses indicated a strong preference for MCQ-based probes and identification of the usefulness of formative assessment, whereas high scores indicated preference opposed to the technique. From these results, it is only possible to confirm a preference towards probes as a method of formative assessment. The questionnaire does not aim to determine other preferred methods.

The questionnaire was designed with objectives of formative assessment in mind, drawn from Black and Wiliam (1998). Student perspectives of this process were investigated. For instance, a student that strongly agrees with the statement 'I think that we should only answer questions when we have learned the whole topic' does not value formative assessment and does not portray the feedback-seeking traits of the growth mindset (Cutumisu, 2019; Wiliam, 2011b). The purpose of each questionnaire statement is outlined in Table 3. The statements were designed to investigate specific aspects of formative assessment (probes) explored in the literature review. Following good research practice guidelines, care was taken to provide statements that did not lead students to select answers

(Bell & Waters, 2014; Munn & Drever, 2004). Clear instructions accompanied the questionnaire (Bell & Waters, 2014).

| Statement | Research purpose of statement |
|---|---|
| Answering questions in science helps me to work out which things I need to work on | Identification of formative assessment usefulness<br>Evidencing intrinsic motivation |
| If I get a question wrong, I know I will understand the correct answer after class | Identification of formative assessment usefulness<br>Evidencing intrinsic motivation |
| Answering questions in science can be enjoyable | Perspectives of formative assessment probes<br>Evidencing intrinsic motivation |
| Multiple choice questions are a good way of showing what I understand | Identification of formative assessment usefulness |
| I think that we should only answer questions when we have learned the whole topic | Identification of formative assessment usefulness<br>Evidencing intrinsic motivation |
| Answering questions can make me feel more confident about what I already know | Evidencing intrinsic motivation |
| Multiple choice questions really make me think | Perspectives of formative assessment probes<br>Evidencing intrinsic motivation |
| My teacher uses questions to understand how much I know about the topic | Identification of formative assessment usefulness |
| I prefer multiple choice questions to open-ended questions | Affective perspectives of assessment probes |
| Usually, I already know stuff about the things we learn in science before lessons | Identification of formative assessment usefulness |
| I worry about getting questions wrong in science | Perspectives of formative assessment probes<br>Evidencing intrinsic motivation |

**Table 3: Questionnaire statements and their research purposes**

Students were then asked to comment upon a statement they felt strongly about, using an open question. This allows an appreciation of the factors important to students and gathers qualitative data regarding opinions that may not be covered by statements. This reflects the responsive nature of action research cycles (McNiff, 2017). Using written responses, future research could investigate areas suggested by students. Post-intervention, the same questionnaire was used to establish changes to student perceptions, with specific instructions to recall the probes when responding.

## Intervention outline

At the start of or during lessons, at least one research-supported formative assessment probe was used. Based on published examples, (Kelley, 2018; UoYSEG, 2020), the stem of the question was contextualised to allow application of knowledge, rather than recital, otherwise identified as a common disadvantage of diagnostic questions (Roberts, 2006). The first tier (an MCQ) offered plausible distractors, and one scientifically correct answer. Following research from Shin, Guo and

Gierl (2019) regarding effective design, distractors were deigned to elicit evidence of known common misconceptions. Based upon recommendations from Wiliam (2011, p.97), responses were not designed to be debatable, and external conditions or factors would not have impacted interpretation, making them appropriate for diagnostic testing. Each probe consisted of an MCQ to tackle recall, and a further MCQ or written answer space to evidence scientific explanation.

Where probes were not self-made following research-led guidelines, questions were provided by the University of York's BEST programme (UoYSEG, 2020). Table 4 outlines the nature of the formative assessment probes used across this scheme of work.

| Lesson and Intervention Title | Misconception or fundamental knowledge being tackled |
|---|---|
| **Lesson 1 – What is a disease?**<br>Interventions:<br>What is a disease?<br><br>Causes of disease | **What is a disease?**<br>Students commonly associate 'health' with only physical fitness and describe health as the absence of physical issues.<br>**Causes of disease**<br>This question tackles similar ideas based on relating 'disease' to physical traits such weight and lifestyle choices such as healthy eating. |
| **Lesson 2 – The spread of disease**<br>Interventions: Catching a cold<br><br>Food poisoning | **Catching a cold**<br>This question tackles misconceptions and old wives' tales associated with the common cold. Distractors are based on these misconceptions.<br>**Food Poisoning**<br>This question targets understanding of how bacterial infection can be spread via food, and the infection process by bacteria. |
| **Lesson 3 – The spread of disease**<br>Intervention: Managing measles | **Managing measles**<br>This MCQ tackles misconceptions about viruses, including that viruses are 'alive'. |
| **Lesson 4 – Non-communicable diseases**<br>Intervention:<br>Can he catch it? | **Can he catch it?**<br>This MCQ clarifies the difference between communicable and non-communicable disease, and checks student recall that heart disease is not infectious. |
| **Lesson 5 – Risk factors and data interpretation**<br>Intervention: Cause and Effect | **Cause and effect**<br>This question tackles misconceptions regarding risk factors such as a perceived 'immunity' to disease due to lifestyle choices. |
| **Lesson 6 – Genetic diseases**<br>Intervention:<br>Passing it on | **Passing it on**<br>Students identify whether diseases are genetic, non-communicable, or communicable. |
| **Lesson 7 – Disease treatment**<br>Interventions:<br>Who needs antibiotics?<br>Will it work? | **Who needs antibiotics?**<br>Students are presented with different patients and decide who needs antibiotics based on their learning. Only one patient has a bacterial infection.<br>**Will it work?**<br>Linked to the previous probe and related to the action of antibiotics and starts to introduce antibiotic resistance. |

**Table 4: Formative assessment probe interventions across lessons**

MCQs are notoriously difficult to design (Brown & Abdulnabi, 2017; Roberts, 2006), and although research was conducted to support the formulation of multiple examples, in the interests of time, BEST resources were also used, providing adequate supporting research could be evidenced. The probes 'Catching a Cold' and 'Managing Measles' are self-made, based on research, whilst the remaining probes taken from BEST resources (Table 4). Appendix One provides a series of tables detailing the interventions used in each lesson, the intervention's source and style, and supporting evidence related to the intervention.

In critique of MCQ use, Paxton (2000) argues that MCQs fail to assess further than surface level knowledge. To investigate this in practice, tier two of the probes consisted of either written explanation prompts or a second MCQ to analyse whether tier one alone identifies misconceptions (in which case both tiers should elicit similar responses), and the ease of assessing scientific explanation via probes. When two four-option MCQs were used, even if the student has guessed the first tier, there is merely a 12.5% chance of being correct for both answers using random guesses. In written explanations, students have freedom to justify their answers, again revealing guesswork. Students willing to admit they had guessed tier one were instructed to do so. The nature of action research is that the intervention may be refined in future cycles (McNiff, 2017). Questions eliciting low success rates could be used as evidence to adjust either the scheme of work, formative assessment probes, or question structure.

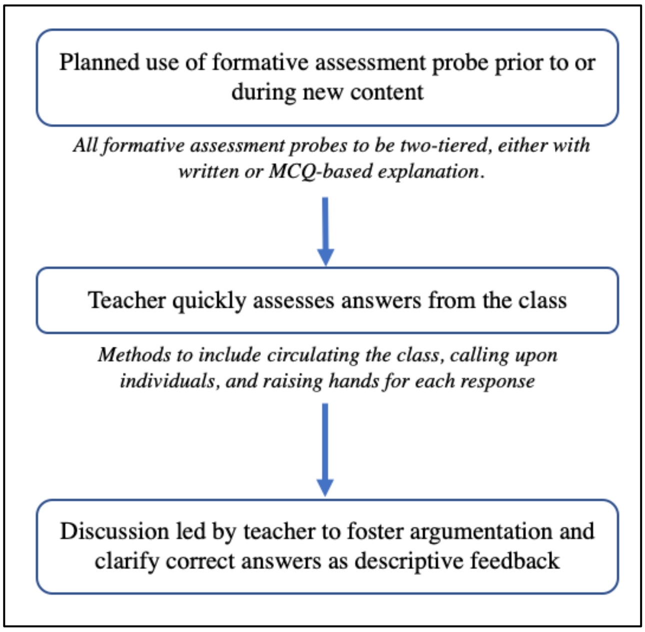


**Figure 3: An outline of the planned intervention, based on research critiqued in the literature review**

Following Keeley (2018), probes were used prior to and during instruction and followed by verbal feedback (Figure 3). Formative assessment probes were reviewed and discussed as a class, with feedback used to encourage argumentation (Bulunuz & Bulunuz, 2013; Keeley, 2013a).

### Ethics

A plan of the proposed study was discussed with subject lecturers of the Faculty of Education (University of Cambridge) and approved. Throughout the study, there was strict adherence to the BERA code of ethics for educational research (BERA, 2018), alongside ethical considerations requested by the placement school. This included distributing a letter to parents/guardians requesting permission for their child to be included in the study, with a correspondence email address given via which students could be removed from the study data by request. Throughout the study, students remained informed, and their anonymity was maintained - probe sheets and questionnaires could not be linked to the respondent (Bell & Waters, 2014, p.51). Research guides were consulted to ensure ethical standards were upheld.

## Findings

The findings of this study are presented with regard to the research questions. Due to small a sample size (n=32), findings are presented in a descriptive manner.

### RQ1: How effective is the use of formative assessment probes in achieving effective formative assessment?

Diagnostic testing is believed to effectively elicit evidence of learning (Wiliam, 2011a). To investigate this as a key feature of effective formative assessment (Black & Wiliam, 1998), research-supported probes were used prior to and during instruction to inform teaching and learning.

When used prior to instruction, the first tier of the probe appeared effective in evidencing misconceptions, but analysis of the second tier revealed this alone is unrepresentative of student learning. For example, the first tier of the probe ‘Food Poisoning’ revealed knowledge was well established as evidenced by the data table presented in Figure 4 showing a 93% (28/30) success rate (shading indicates correct responses). However, scientific explanations to support this were lacking, evidenced by a spread of tier two results (see the second column of the data table for specific numerical results) that indicate underlying misconceptions. The two accompanying radar graphs (also in Figure 4) further illustrate the results from the ‘Food Poisoning’ probe, revealing good retention of factual knowledge, but a spread of answers across misconceptions regarding scientific explanations. This provided useful information regarding the nature of feedback required. The results also emphasise the importance of incorporating opportunities for explanations. Reflecting disadvantages of MCQs anticipated by Bulunuz et al. (2016), the first tier alone is too discrete and specific in the information it targets, and as a result, fails to identify misconceptions.

| Answer option | Food Poisoning – Tier 1 | Food Poisoning – Tier 2 |
|---|---|---|
| A | 1 | 4 |
| B | 28 | 2 |
| C | 1 | 18 |
| D | 0 | 6 |
| Total | 30 | 30 |

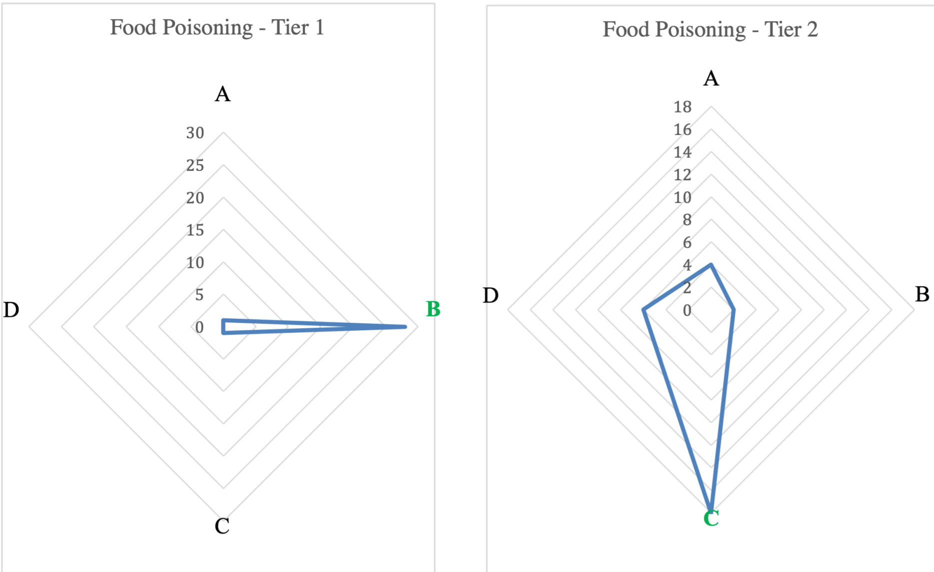


**Figure 4: The results of the probe 'Food poisoning': data table & radar graphs**

When probes consisted of two MCQs, feedback was guided towards prevalent misconceptions. Using the probe 'Causes of Disease' (see Figure 5 for data table and radar graphs, again table shading indicates correct responses), 24 out of 27 students identified the correct use of the colloquial term 'germs' but struggled to then use the scientific term 'microorganisms' in a follow up question, with 14 students selecting incorrect options. This information was quickly accessible by asking students to raise their hands for each option. This was used to guide argumentation to tackle a prevalent misconception (that 'germs' are not microorganisms). This enhances teaching as the teacher knows that generalised feedback will be useful to nearly half the class.

| Answer option | Causes of Disease – Tier 1 | Causes of Disease – Tier 2 |
|---|---|---|
| A | 1 | 3 |
| B | 24 | 13 |
| C | 0 | 0 |
| D | 2 | 11 |
| Total | 27 | 27 |

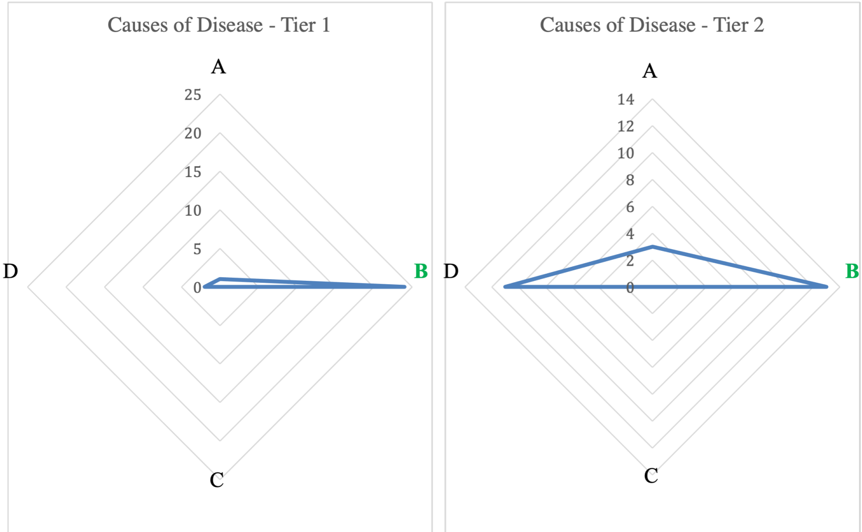


**Figure 5: The results of formative assessment probe 'Causes of Disease': data table and radar graphs**

Conversely, when written answers were used as the second tier, time required to access explanations halted feedback, as false positives were created when students guessed the correct answer for the initial MCQ. When guessing, students were encouraged to write down that they had guessed. Without time to investigate written answers and account for guesswork, probes obtaining high tier one success rates incorrectly portrayed that the class needed no further support. For example, for the probe 'Cause and Effect', 80% of students selected the correct answer for tier one (Table 5 - shaded cells indicate correct responses).

| Cause and Effect - Tier 1 | |
|---|---|
| Answer option | Responses |
| A | 0 |
| B | 6 |
| C | 24 |
| D | 0 |
| Total | 30 |

**Table 5: The results from the first tier of formative assessment probe 'Cause and Effect'**

Without time in lesson to review individual explanations, this was the only accessible evidence of learning. When explanations were reviewed, 50% of students providing ‘correct’ answers indicated they had guessed or could not give relevant explanations. This does not include incorrect tier one answers justified by misconceptions. The prevalence of this issue across formative assessment probes with written answer explanation prompts raises questions regarding their accuracy in evidencing student learning for real-time response (Table 6).

| Percentage of correct tier one answers produced by means of guesswork or based on no relevant explanation | | | |
|---|---|---|---|
| Cause and Effect | Managing measles | Catching a Cold | Defining Health |
| 50% | 30% | 39% | 20% |

**Table 6: The percentage correct tier one answers that were guessed or unjustified in tier two written explanations**

When used during instruction to inform teaching, formative assessment probes were effective in evidencing strengths of the class. This was illustrated by probes achieving 100% success rates. The probe ‘Can they catch it?’ was used to evidence learning of the difference between communicable and non-communicable diseases during instruction (Table 7 – shaded cells indicate correct responses). All students were able to select the correct factual information and scientific explanation. This enhances teaching by highlighting that new content can now be introduced, moving the narrative forwards only once students can recall and explain content learned thus far.

| Answer option | Can they catch it?<br>Tier 1 (recall) | Can they catch it?<br>Tier 2 (explanation) |
|---|---|---|
| A | 31 | 0 |
| B | 0 | 0 |
| C | 0 | 31 |
| D | 0 | 0 |
| Total | 31 | 31 |

**Table 7: The results of formative assessment probe ‘Can they catch it?’**

*Attainment data*

At the end of the topic, students completed an online assessment of content covered. The mean percentage achieved on the test compared to that of the previous topic was improved by 12% across both foundation and higher-level tests. Figure 6 presents a box plot graph comparing student attainment in this and the previous summative test. Figure 7 presents a bar chart illustrating the attainment results for each student in the two tests. This increase in attainment may reflect the incorporation of an effective formative assessment technique, though further research would be required to clarify this. Black and Wiliam (1998) reported typical size effects of 0.4-0.7 for formative assessment experiments, implying that effective formative assessment can entail significant learning gains, supporting this link between the use of formative assessment probes and increased attainment in summative assessment.

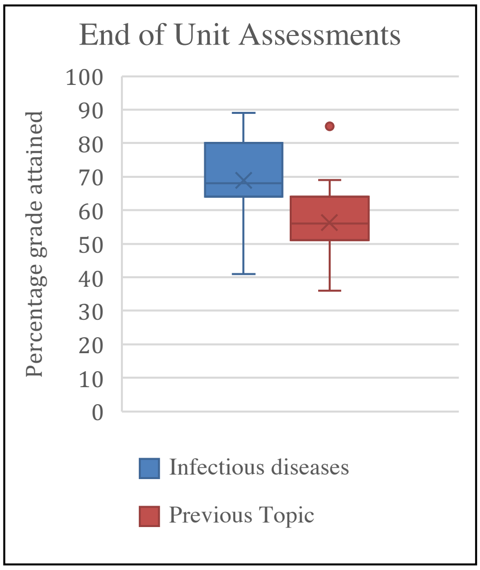


**Figure 6: Student attainment in summative tests**

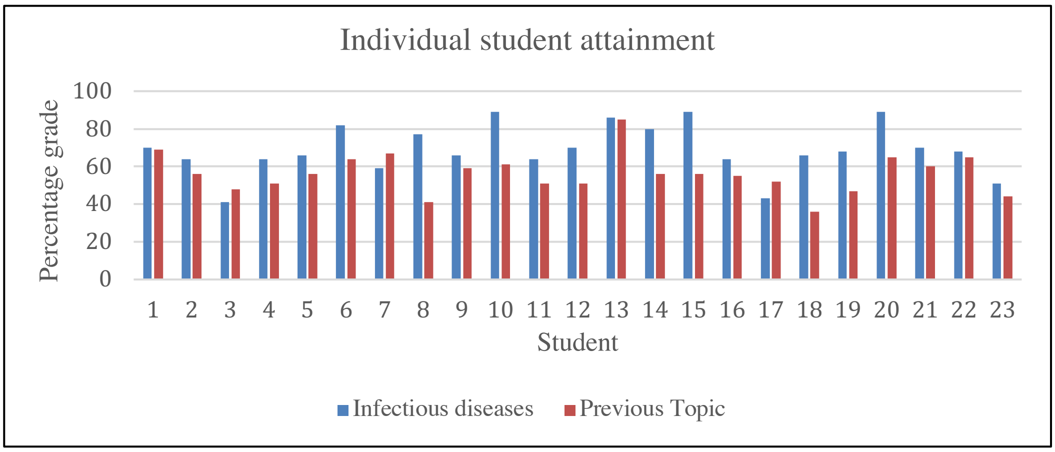


**Figure 7: Individual student attainment in summative tests**

As shown in Figure 7, only one student (student seven) scored higher in the previous summative test, with the most significant progress shown by student eight, with a 36% improvement upon their previous score.

*Classroom Observations*

During discussion following the observation, teachers commented upon the use of probes to foster argumentation and discussions provoking intelligent questions. Observers perceived that formative assessment probes worked well to offer the teacher useful evidence of learning. The majority of students readily engaged with the intervention.

Reflecting upon practice, the researcher noted that probes appeared to encourage responsive teaching. Discussions focused upon specific misconceptions, allowing planning of responses before lessons. However, the inclusion of this tool considerably increased lesson planning time. Time must be taken to pre-identify objectives, collect probes that reflect these, and plan teaching in response to multiple outcomes. Conversely, formative assessment probes were typically short activities, lasting approximately ten minutes unless prevalent misconceptions were detected.

**RQ2: How are student's perceptions of formative assessment probes affected by their consistent use throughout a scheme of work?**

*Pre-intervention questionnaire*

The findings of the initial questionnaire revealed a preference towards the MCQ style used in formative assessment probes (indicated by statement scores of < 3). However, as predicted by Van der Kleij (2019), preferences were not based upon the usefulness of formative assessment. Twenty-eight students provided full responses to the questionnaire (n = 28, female = 15, male = 13), with mean scores for each statement shown in Table 8.

| Statement | Mean Score |
|---|---|
| Answering questions in science helps me to work out which things I need to work on | 3.07 |
| If I get a question wrong, I know I will understand the correct answer after class | 2.52 |
| Answering questions in science can be enjoyable | 3.93 |
| Multiple choice questions are a good way of showing what I understand | 2.03 |
| I think that we should only answer questions when we have learned the whole topic | 3.03 |
| Answering questions can make me feel more confident about what I already know | 2.41 |
| Multiple choice questions really make me think | 2.76 |
| My teacher uses questions to understand how much I know about the topic | 2.55 |
| I prefer multiple choice questions to open-ended questions | 1.55 |
| Usually, I already know stuff about the things we learn in science before lessons | 3.17 |
| I worry about getting questions wrong in science | 3.17 |
| TOTAL | 30.19 |

**Table 8: Mean scores for student responses to pre-intervention questionnaire**

Student responses revealed a preference for the style of formative assessment probe questions, but the formative importance of this was not realised.

Students then selected one question they felt strongly about to comment upon. Inductive coding was used to analyse these responses, with common words, themes or phrases recorded across answers. Initial student comments were focused upon the practical use of probes, rather than the underlying reasons for completing them. Students identified that they preferred MCQ styles, based on ideas that they would be 'easy', 'quick' or offer the opportunity for guesswork. The time taken to complete questions was important to students, possibly as a reflection the high score obtained for 'Answering questions in science can be enjoyable', which suggests a general lack of motivation to engage with formative assessment.

The most commented upon statement was statement nine, regarding a preference for MCQ styles over open-ended questions. Students naturally compared MCQs to other question styles in their answers. Responses were used to create word clouds, with the most frequently used terms/themes appearing in larger text. Two analyses were conducted, one regarding perceptions of open-ended question techniques, and the other regarding perceptions of MCQs (Figures 8 and 9 respectively). This was considered due to the frequent use of both styles in formative assessment probes.

Students who felt they preferred MCQs emphasised the amount of time taken to complete other question styles as a key issue. Students who felt they preferred MCQs to open-ended questions said they preferred MCQs due to guidance from distractors and a belief that they were quicker and easier. Ability to guess answers was viewed positively.

| Word or Theme | Frequency |
|---|---|
| Boring | 6 |
| Too much to write | 4 |
| Too much time | 4 |
| Mistakes | 1 |
| Stressed | 1 |
| Clueless | 1 |
| Unenjoyable | 1 |

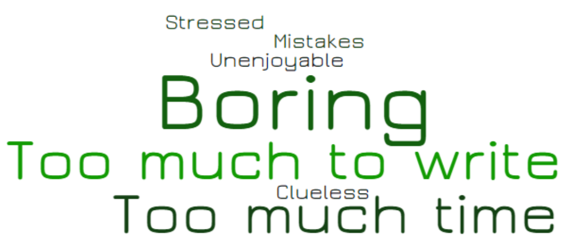


**Figure 8: Student perceptions of open-ended questions**

| Word or theme | Frequency |
|---|---|
| Fun | 1 |
| Quick | 4 |
| Easy | 5 |
| Less Writing | 2 |
| Guess | 2 |
| Helpful choices | 6 |
| Lack explanation | 1 |
| Makes you think | 3 |
| Prefer | 3 |

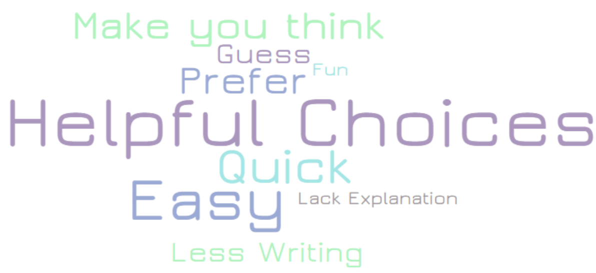


**Figure 9: Student perceptions of MCQs**

Themes reveal that students are concerned with time taken to complete questions and success rate. This was also reflected in student comments. Below, student two identifies a chance of getting the question right by guessing, portraying extrinsic motivation driven by a desire to appear competent, regardless of actual understanding (Ryan & Deci, 2000). Research suggests these behaviours are presented to protect self-worth – appearing competent becomes more important than an intrinsically motivated pursuit of knowledge (Seifert, 2004).

> Student 1: Answering written questions is always boring because there's always too much to write and multiple-choice questions are fun because you only have to tick a quick box and not write an entire paragraph about it
>
> Student 2: Multiple choice are usually quite easy and if you don't know there is still a 50% or 25% chance of getting it
>
> Student 3: If its multiple choice you can't explain why you think it's that or you could have taken a complete guess, so you don't know if you need to work on that question
>
> Student 4: I prefer multiple choice questions because written questions are longer than multiple choice. I would prefer the questions to be multiple choice! Written questions are boring. Really unenjoyable.

Conversely, student three's answer identifies a potential flaw of using MCQs, previously identified in the literature review (Bulunuz et al., 2016); MCQs target discrete, specific knowledge, without explanation. This is interesting as it suggests not only that some students identify flaws in MCQ use, but that students are unfamiliar with the structure of formative assessment probes that elicit evidence of scientific explanation via MCQs. This is a new approach for students, who may have more commonly experienced MCQs used for summative purposes, rather than formative, as previously identified by Roberts (2006).

Students one and four flagged issues with perceived laborious methods, in preference of assessment they believed to provide faster results and feedback. Although the students had not used formative assessment probes, they perceived that MCQs would provide feedback in far less time. Indeed, literature identifies this efficiency as a factor that makes MCQ styles so appealing (Brown & Abdulnabi, 2017; Gierl et al., 2017; Shin, Guo & Gierl, 2019).

*Post-intervention questionnaire*

The mean score (see Table 9) for the questionnaire was reduced post-intervention, indicating a stronger preference for the style of formative assessment probes and an improved understanding of their usefulness. Answers from both Likert ratings and written comments suggested students had progressed towards a growth mindset and intrinsic motivation.

| Statement | Mean Score |
|---|---|
| Answering questions in science helps me to work out which things I need to work on | 2.11 |
| If I get a question wrong, I know I will understand the correct answer after class | 2.18 |
| Answering questions in science can be enjoyable | 2.89 |
| Multiple choice questions are a good way of showing what I understand | 2.00 |
| I think that we should only answer questions when we have learned the whole topic | 2.71 |
| Answering questions can make me feel more confident about what I already know | 2.00 |
| Multiple choice questions really make me think | 2.75 |
| My teacher uses questions to understand how much I know about the topic | 2.04 |
| I prefer multiple choice questions to open-ended questions | 1.32 |
| Usually, I already know stuff about the things we learn in science before lessons | 3.25 |
| I worry about getting questions wrong in science | 2.79 |
| TOTAL | 26.04 |

**Table 9: Student responses to post-intervention questionnaire**

The lower overall mean score of 26.04 (n = 28) seems to indicate that students are more aligned with the growth mindset and feedback-seeking preferences than before the intervention (pre-intervention mean - 30.19, n = 28). Thus student responses revealed a preference for the intervention style, and the formative importance of this was understood by students. Indeed, in contrast to earlier comments reflecting extrinsic motivators, post-intervention comments recognised the usefulness of probes in aiding personal learning. Below, students 1 and 4 note the role of formative assessment for the teacher in continuously assessing understanding to shape teaching. Both students appear to understand that the identification of misconceptions allows the teacher to support students. Further, student 2 identifies that assessment should be continuous to tackle misconceptions as they arise, as suggested by Black and Wiliam (1998).

> Student 1: It is good that [the teacher] checks it and it helps us improve
>
> Student 2: I disagree with only doing questions after we have finished the topic because they help you understand as you go along, and they check that you know the right things.
>
> Student 3: I strongly disagree with worrying about getting a question wrong in science because when you go over it and mark it you then learn the right answer.
>
> Student 4: I think the multiple-choice questions are very helpful because it gives the teacher a chance to see what we need to work on to improve.
>
> Student 5: I think multiple choice questions are good because it tests your knowledge and understanding on the subject.

In contrast to pre-intervention comments that did not identify the educational purpose of formative questioning, student 5 correctly identifies the purpose of the two separate tiers of formative assessment probes, using language appropriate for describing the learning journey. This suggests the student self-identifies the usefulness of feedback provided by formative assessment probes, which research identifies as key to impacting achievement (Harks et al., 2014). This may be due to recognition of a formative rather than summative purpose of diagnostic testing, which research suggests may have been lacking prior to intervention (Roberts, 2006).

Furthermore, Student 3's comment reflects the resilience necessitated by the growth mindset – accepting, appreciating, and working upon feedback, which contributes to learning (Black & Wiliam, 1998; Cutumisu, 2019; Wiliam, 2011b). This student's comments reflect an acceptance of the formative purpose of the probes as integral to learning. Furthermore, feedback that is corrective and clarifies the school science narrative is preferable (Attali, Laitusis & Stone, 2016), and the student identifies that probes achieve this. Across the class, the mean score for the related statement (statement 11) fell from 3.17 to 2.79, suggesting a slight overall adjustment towards positive perceptions of formative assessment following the intervention.

*Classroom Observations*

Guided by research developed by Skinner et al. (2009), a checklist of identifiers of student engagement and disaffection further investigated student perception of and engagement with probes. Results suggested lower ability students were more responsive to the tool and identified a danger of students becoming distracted if probes were not demanding enough. One higher attaining student completed the probe quickly but then turned to talk to classmates, suggesting a lack of cognitive demand for higher attaining students. Observations suggested the majority of students were happy to engage, with a minority appearing 'bored'. In utilising a checklist of previously researched factors related to motivation and engagement (Skinner et al., 2009), observations generally supported positive student perceptions of probes evidenced by questionnaires.

## Discussion

### RQ1: How effective is the use of formative assessment probes in achieving effective formative assessment?

Following the literature review, effective formative assessment was understood to accurately evidence student learning and provide feedback that moves learning forwards (Black & Wiliam, 1998). In this study, formative assessment probes have been shown to elicit evidence of learning to tackle common misconceptions via verbal feedback in secondary science lessons. Attainment in summative testing was raised by 12% for both 'foundation' and 'higher' level students, suggesting the intervention enhanced learning.

Results revealed the strengths of the technique in evidencing learning, specifically the identification of misconceptions. Identification of misconceptions via probes to guide teaching potentially raised attainment in summative tests, supporting the work of Driver et al. (2015) and Galvin et al. (2015) who highlight misconceptions as an important issue for science education. By basing diagnostic tools upon existing literature regarding common misconceptions, the teacher can plan appropriate responses ahead of lessons to ensure targeted feedback. This entails a cycle of misconception recognition, reduction, and removal, as identified by Galvin et al. (2015).

It was clear that planning is an absolute requirement of effectively using this pedagogical tool. In effect, the teacher must implement a backwards planning strategy when using formative assessment probes, identifying learning objectives and arising misconceptions prior to lessons. This backwards strategy has been recommended in the literature to ensure effective use of formative assessment probes (Whitehouse, 2014). However, throughout the study it was noted that although the use of probes may only require ten minutes of lesson time, the planning process far exceeds this, especially if teachers devise their own probes. The time-consuming process could deter teachers from using the technique.

During the planning process, the structure of formative assessment probes should be considered. In this study, the results of the two tiers of probes was often dissimilar, with students appearing to have learnt concepts, but explanations revealing that this learning was restricted to factual, discrete information, as limitations outlined by Bulunuz et al. (2016) had previously anticipated. To counter this, results suggest the second tier is important in achieving evidence that is representative of student learning. This should be appreciated when teachers use diagnostic testing to drive formative assessment.

Furthermore, probes consisting of two MCQs, rather than an MCQ and a written explanation, were more suitable for rapidly identifying misconceptions to be addressed during the lesson. The time taken to read or verbally request written answers was not freely available, and if assessment of prior learning is to shape teaching within the same lesson, assessment must be rapid. Immediate feedback has been found to be more effective than delayed feedback in aiding complex learning (Attali, Laitusis & Stone, 2016), further emphasising the importance of a design that allows this.

Additionally, it must be noted that only argumentation was implemented throughout this intervention. Augmentation was chosen as a feedback tool based on existing research (Bulunuz & Bulunuz, 2013; Keeley, 2013a; Keeley, 2016), however, as discussed by Wiliam (2011), there are a wealth of strategies available to give useful feedback from formative tasks, which this study cannot comment upon. Future research could investigate the use of various feedback strategies (for example, Keeley, 2015).

**RQ2: How are student's perceptions of formative assessment probes affected by their consistent use throughout a scheme of work?**

Questionnaire responses suggested that post-intervention, students more readily identified the usefulness of formative assessment. In support of existing research, analysis of student responses to open questioning also revealed evidence of intrinsic motivation post-intervention, thought to be far more effective in promoting engagement the than extrinsic motivational styles evidenced pre-intervention (Ryan & Deci, 2000). Classroom observations indicated that students engaged well with probes, although lower attaining students potentially benefited to a greater extent in terms of engagement. Due to a small sample (n=6), further research is required to support this.

It is important that students self-identify the usefulness of formative assessment to ensure it positively impacts attainment (Harks et al., 2014). Results from this study suggest that use of formative assessment probes aids student identification of the usefulness of formative assessment. Post-intervention questionnaire responses revealed that student perceptions of formative assessment probe use were largely positive (indicated by mean scores of < 3). Evidence of intrinsic motivation styles and increased attainment also followed the intervention, supporting the links between positive student perceptions, motivation, and performance proposed in Van der Kleij's (2019) research. Although these results are promising, this research lasted just eight lessons and long-term effects cannot be predicted. Long-term effects must be considered, as student perceptions of being 'monitored' can reportedly have negative impacts upon achievement (Kyaruzi et al., 2019). Perceptions may change over time, and teachers should be wary of this. The impact of long-term use upon student perceptions cannot be predicted by this study.

In the short-term, questionnaire responses and observations imply that these students enjoyed using formative assessment probes. Mean scores from Likert ratings indicated that perceptions of formative assessment were improved by continuous use of the tool (for example, pre-intervention mean score for statement 3 = 3.93; Post-intervention = 2.89), and the majority of students appeared happy when completing probes. It is generally accepted that positive classroom environments are preferable and that children learn best when educated in positive, supportive environments, providing relevant content and effective feedback (Young, 2014). The results of this study reflect these conditions. Furthermore, when students are willing to engage with formative assessment, the benefits previously evidenced can be realised (Wiliam, 2011b). As student perceptions have been evidenced to impact motivation to engage (Van der Kleij, 2019), the positive perceptions evidenced in this study imply that the technique has an increased likelihood of enhancing learning.

Therefore, the potential impact of these results should not be underestimated, especially as analysis of qualitative data via inductive coding revealed that students perceive other questioning styles as time-consuming, unenjoyable, or boring. Young (2014) suggests that improved academic results may be linked to the inclusion of techniques that promote positive student views – a finding supported by this study, with implications for future practice regarding the identification and utilisation of techniques that children enjoy, to enhance learning.

**Limitations**

Here, limitations specifically related to the methodology of the study are discussed. Key limitations included use of questionnaires, a lack of student interviews, and the absence of inferential statistics.

Qualitative data collected via questionnaires regarding student perceptions was useful in supporting Likert ratings. However, it may have been useful to conduct follow-up interviews, allowing students to clarify comments. The use of questionnaires presents limitations. As discussed by Munn and Drever (2004, p.5-6), the data collected from questionnaires is descriptive; it does not offer explanation. For example, post-intervention, 32% of students (total n = 28) indicated they were 'not sure' whether answering questions in science could be enjoyable. Only one student commented upon the basis of this selection, and the explanations of eight students remain undisclosed. Questionnaires must use closed questioning to aid the descriptive, 'fact-finding' nature of results (Wellington, 2000, p.102), but this is at the expense of explanatory data (Munn & Drever, 2004, p.5). Questionnaire responses could have been strengthened by interviews, evidencing the thoughts and emotions of participants which are often not accessible via other methods (Wellington, 2000, p.71). Unfortunately, opportunity to collect this data was restricted by school closures. Its inclusion would have provided qualitative data to conduct more substantial content analysis.

Concerning the intervention itself, teacher effect may have influenced results. As a pre-service teacher, subject knowledge may have been less robust than that of an experienced teacher, which Sadler and Sonnert (2016) suggest can impact the effectiveness of formative assessment probes. However, this action research project was designed to provide insight into the practice of one classroom (McNiff, 2017). Although this entails low generalisability, action research is commonly conducted by educators to inform personal practice, offering only a useful insight to other practitioners (McNiff, 2017, p.41). More experienced practitioners may achieve further enhancement of teaching and learning, aided by robust subject knowledge – a prediction based on existing literature that requires further research.

As a further result of the research design, the sample size does not necessitate the use of inferential statistics. They would not benefit the study of this limited population (Wilson, 2012). Reflecting the small scale of this research, convenience sampling on a non-probability basis was used (Wellington, 2000, p.59). In doing so, and due to the nature of educational research, results are reflective of specific actions and consequences, limiting generalisability (Biesta, 2010). To draw more confident conclusions, larger samples are required, capturing the perspectives of a wider diversity of students and necessitating inferential statistics.

## Conclusions and future research

In conclusion, formative assessment probes have been evidenced to elicit useful evidence of student learning, and results from student questionnaires suggested continuous use of the technique increased positive student perceptions and perceived usefulness of formative assessment, reflective of intrinsic motivation styles. Benefits included rapid assessment and response when questions were driven by misconceptions pre-determined by the literature. This allows teachers to anticipate feedback directions. Limitations of the intervention included an inability to assess student written answer explanations during lessons, reducing power to guide learning. In effect, this may restrict questions to a diagnostic (rather than formative) purpose. Furthermore, the first tier of probes alone is often not sufficient to evidence learning, and a second tier regarding scientific explanation is necessary to uncover misconceptions.

| General recommendations | |
|---|---|
| 1 | When used as formative assessment probes, diagnostic questions should be based upon existing research regarding common misconceptions to elicit evidence of these issues and to aid planning of feedback |
| 2 | Practitioners should consider the design of formative assessment probes to ensure they reflect the timing of feedback. When used during lessons, formative assessment probes are most useful when they consist of two MCQs. This evidences knowledge and scientific reasoning in a quick, accessible format to guide the lesson in real time. Written answer explanations do not offer the same rapid feedback response. |
| 3 | The consistent use of formative assessment probes may improve student engagement with the usefulness of formative assessment. Research suggests this is important for effective formative assessment as the learning lies with the student. Teachers should utilise tools such as formative assessment probes that encourage this. |

**Table 10: General recommendations for practice**

Future research cycles should reflect the findings of this work, and general recommendations for practice arising from this study are outlined in Table 10. As student responses and quantitative data suggest that MCQ questions are more effective and positively received in practice, future cycles could specifically investigate the use of probes consisting of two MCQs, removing written explanations. Results here suggest this could further facilitate rapid feedback.

Alternatively, future cycles could investigate the long-term effects of using formative assessment probes upon teaching and learning. This study cannot clarify whether positive student perceptions will be upheld, especially in light of research from Kyaruzi et al. (2019), who suggest that being 'monitored' can have negative impacts upon learning.

Finally, future cycles could identify whether probes are more effective in supporting learning in certain student groups. Observations began to suggest that lower attaining students may engage with the probes to a greater extent than higher attaining students, who conversely completed the tasks efficiently and did not push themselves to provide further explanation. A research cycle focused upon this aspect of formative assessment probe use could clarify this hypothesis with larger sample sizes.

## References

Andersson, C. & Palm, T. (2018). Reasons for teachers' successful development of a formative assessment practice through professional development – a motivation perspective, *Assessment in Education: Principles, Policy & Practice*, *25*(6), 576-597.

Attali, Y., Laitusis, C., & Stone, E. (2016). Differences in Reaction to Immediate Feedback and Opportunity to Revise Answers for Multiple-Choice and Open-Ended Questions. *Educational and Psychological Measurement*, *76*(5), 787–802.

Bell, J. & Waters, S. (2014). *Doing your research project: a guide for first-time researchers.* (6th ed.). Berkshire: Open University Press

BERA (2018). Ethical Guidelines for Educational Research, 4th edition. Retrieved from: https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018-online

Biesta, G. J. J. (2010). *Good education in an age of measurement*. London: Routledge.

Black, P. & Wiliam, D. (1998) *Inside the Black Box.* Phi Delta Kappan, Bloomington, *80*(2).

Brown, G.T.L., & Abdulnabi, H.H.A. (2017). Evaluating the Quality of Higher Education Instructor-Constructed Multiple-Choice Tests: Impact on Student Grades. *Frontiers in Education, 2*(24).

Bulunuz, N., Bulunuz, M., Karagoz, F., & Tavsanli, O.F. (2016). Achievement Levels of Middle School Students in the Standardized Science and Technology Exam and Formative Assessment Probes: A Comparative Study. *Journal of Education in Science, Environment and Health*, *2*(1).

Bulunuz, M., & Bulunuz, N. (2013). Formative Assessment in Science Teaching and Demonstration of its Effective Implementation. *Journal of Turkish Science Education*, *10*(4).

Cutumisu, M. (2019). The association between feedback-seeking and performance is moderated by growth mindset in a digital assessment game. *Computers in Human Behavior, 93*, 267-278

Dokter, E., Pompea, S., Sparks, R., & Walker, C. (2010). The development of formative assessment probes for optics education. *Proceedings of SPIE, 7783*(1), 778309-778308.

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (2015). *Making Sense of Secondary Science*. London: Routledge.

Duschl, R.A., & Osborne, J. (2002). Supporting and Promoting Argumentation Discourse in Science Education. *Studies in Science Education*, *38*, 39-72.

Exampro (2020). The essential exam preparation tool for teachers. Retrieved from: https://www.exampro.co.uk/science/

Fuchs, T.T., & Arsenault, M. (2017). Using test data to find misconceptions in secondary science. *School Science Review, 98*(364).

Hanson, M. & Gluckman, P. (2011). Developmental origins of noncommunicable disease: population and public health implications. *The American Journal of Clinical Nutrition*, *94*, 1754S-1758S.

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology*, *34*, 269e290.

Hirst, P. (1975). What is teaching? In R. S. Peters (ed.) *The Philosophy of Education*. London: Routledge and Kegan Paul.

Galvin, E., Simmie, G.M., & O'Grady, A. (2015). Identification of misconceptions in the teaching of biology: a pedagogical cycle of recognition, reduction and removal. *Higher education of social science, 8*(2), 1-8.

Gierl, M., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, Analyzing, and Using Distractors for Multiple-Choice Tests in Education: A Comprehensive Review. *Review of Educational Research, 87*(6), 1082-1116.

Johnson, M., & Bungum, T. (2013). Identifying and Reconstructing Common Cold Misconceptions Among Developing K-12 Educators. *American Journal of Health Education*, *44*(3), 169-175.

Keeley, P., Eberle, F., & Farrin, L. (2005). *Uncovering Student Ideas in Science, Volume 1: 25 Formative Assessment Probes*. Arlington :National Science Teachers Association Press

Keeley, P. (2013a). Is It a Solid? Claim Cards and Argumentation. *Science and Children, 50*(9), 26-28.

Keeley, P. (2013b). Formative Assessment Probes: Labeling versus Explaining. *Science and Children, 50*(6), 32-34.

Keeley, P. (2015) Constructing Cl-Ev-R explanations to formative assessment probes. *Science and children*, 53(3), p.26

Keeley, P. (2016). Formative Assessment Probes: Talk Moves. A Formative Assessment Strategy for Fostering Productive Probe Discussions. *Science and Children, 53*(8), 24.

Keeley, P. (2018). *Uncovering Student Ideas in Science, Volume 1: 25 Formative Assessment Probes.* (2nd ed.) Arlington: National Science Teachers Association.

Kilgour, L., Matthews, N., Christian, P., & Shire, J. (2015). Health literacy in schools: prioritising health and well-being issues through the curriculum. *Sport, Education and Society*, *20*(4), 485-500.

Kyaruzi, F., Strijbos, J-W, Ufer S., & Brown, G.T.L. (2019). Students' formative assessment perceptions, feedback use and mathematics performance in secondary schools in Tanzania. *Assessment in Education: Principles, Policy & Practice*, *26*(3), 278-302.

McNiff, J. (2017). *Action Research: All you need to know*. London: Sage Publishing.

Millar, R. (2016). Using assessment to drive the development of teaching-learning sequences. Retrieved from: https://www.york.ac.uk/media/educationalstudies/documents/research/uyseg/diagnosticassessment/2016%20Millar%20-%20ESERA%20e-Proceedings%20paper%20on%20YS%20project.pdf

Millar, R. & Hames, V. (2003). Using Diagnostic Assessment to Enhance Teaching and Learning. A Study of the Impact of Research-informed Teaching Materials on Science Teachers' Practices. Retrieved from: https://www.york.ac.uk/media/educationalstudies/images/curriculumprojects/c21/EPSE%20P1%20final%20report.pdf

Munn, P., & Drever, E. (2004). *Using Questionnaires in Small-Scale Research*. The SCRE Centre, University of Glasgow. Edinburgh: Luton Press.

Paxton, M. (2000). A Linguistic Perspective on Multiple Choice Questioning. *Assessment & Evaluation in Higher Education*, *25*(2), 109-119.

Pelikan, J. M., Ganahl, K. & Roethlin, F. (2018). Health literacy as a determinant, mediator and/or moderator of health: empirical models using the European Health Literacy Survey dataset. *Global Health Promotion*, *25*(4), 57-66.

Queloz, A.C., Klymkowsky, M.W., Stern, E., Hafen, E., & Köhler, K. (2017). Diagnostic of students' misconceptions using the Biological Concepts Instrument (BCI): A method for conducting an educational needs assessment. *PLoS ONE, 12*(5), E0176906.

Roberts, T. (2006). The Use of multiple choice test for formative and summative assessment. *Conferences in Research and Practice in Information Technology Series. 52*, 175-180. Ryan, R.M., and Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Sadler, P.M, & Sonnert, G. (2016). Understanding misconceptions Teaching and learning in middle school physical science. *American Educator*, *6*(7).

Shaw, K., Horne, K., Zhang, H., & Boughman, J. (2008). Essay Contest Reveals Misconceptions of High School Students in Genetics Content. *Genetics, 178*(3), 1157-68.

Shin, J., Guo, Q., & Gierl, M.J. (2019). Multiple-Choice Item Distractor Development Using Topic Modeling Approaches. *Frontiers in Psychology, 10*(825).

Seifert, T. (2004) Understanding student motivation. *Educational Research*, *46*(2), 137-149.

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A Motivational Perspective on Engagement and Disaffection: Conceptualization and Assessment of Children's Behavioral and Emotional Participation in Academic Activities in the Classroom. *Educational and Psychological Measurement*, *69*(3), 493–525.

Slaughter, V., & Ting, C. (2010). Development of ideas about food and nutrition from preschool to university. *Appetite, 55*(3), 556-564.

Trauth-Nare, A., & Buck, G. (2011). Assessment FOR Learning. *The Science Teacher, 78*(1), 34-39.

University of York Science Education Group (UoYSEG). (2020). Best Evidence Science Teaching - The best teaching draws on the best evidence. Retrieved from: https://www.stem.org.uk/best-evidence-science-teaching

Van der Kleij, F., Vermeulen, J., Schildkamp, K., & Eggen, T. (2015). Integrating data-based decision making, Assessment for Learning and diagnostic testing in formative assessment. *Assessment in Education: Principles, Policy & Practice, 22*(3), 324-343.

Van der Kleij, F.M. (2019). Comparison of teacher and student perceptions of formative assessment feedback practices and association with individual student characteristics. *Teaching and Teacher Education, 85*, 175-189.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Wellington, J. (2000). *Educational Research: Contemporary issues and practical approaches*. London: Continuum.

Whitehouse, M. (2013). Embedding Assessment to Improve Learning. *School Science Review, 95*(351), 52-56.

Whitehouse, M. (2014). Using a Backward Design Approach to Embed Assessment in Teaching. *School Science Review, 95*(352), 99-104.

Wiliam, D. (2011a). *Embedded Formative Assessment.* Bloomington: Solution Tree Press.

Wiliam, D. (2011b). *Improving learning in Science with Formative Assessment.* Retrieved from: https://s3.amazonaws.com/nstacontent/PB219X-1.pdf?AWSAccessKeyId=AKIAIMRSQAV7P6X4QIKQ&Expires=1575799363&Signature=wvZJnqurp3q1N9l%2bZxOM%2bBFYNQo%3d

Wilson, E. (Ed.). (2012). *School-based research: a guide for education pupils*. London: Sage.

Wind, S.A. & Gale, J.D. (2015). Diagnostic Opportunities Using Rasch Measurement in the Context of a Misconceptions-Based Physical Science Assessment. *Science Education*, *99*(4), 721-741.

Yekeen, A.F.A., Bell, J., Boorman, A., Hancock C., Kelly, A., Kelly, E., Mulrooney, H.M., &

Young, J. (2014) *Encouragement in the classroom: How Do I Help Students Stay Positive and Focused?* Alexandria, United States: ASCD.

# Appendix One

## Interventions used in each lesson, the intervention's source and style, and supporting evidence related to the intervention

Lesson 1: What is a disease?

| Intervention Title | Misconception or fundamental knowledge being tackled | Source | Style | Supporting research |
|---|---|---|---|---|
| What is a disease?<br><br><br><br><br><br><br><br><br>Causes of disease | **What is a disease?**<br>Students commonly associate 'health' with only physical fitness and describe health as the absence of physical issues.<br><br>**Cause of disease**<br>This question tackles similar ideas based on relating 'disease' to physical traits such weight and lifestyle choices such as healthy eating. 'Germ' and 'bug' are colloquial terms commonly misused in science classrooms. | Modified BEST resources provided by the University of York. Original versions did not allow for student explanation and were modified to enable this. | Two-tier formative assessment probes (MC Q and written explanation) | Berk, 2000: Understanding of health and disease change with developmental stage, with younger students commonly defining health in terms of physical strength<br>Reeve and Bell, 2009: In the US, >60% of children aged 9-11 defined health based upon disease, and linked this to ideas of weight and diet<br>Wang, Zou, Gifford and Dalal, 2014: Students relate being able to physically *do* things to health in general.<br>Maxted, 1984: 12-13-year olds in England used the terms 'bug' and 'germ' instead of 'microbes' or 'microorganisms' in science classrooms.<br>Çetin, Ozarslan, Isik and Eser, 2013: Children aged 14-15 in Turkey associated health and disease strongly with physical activity, weight, and food<br>See also: Brindal, Hendrie, Thompson and Blunden, 2012: for the similar views of primary school children |

Lesson 2: The spread of disease

| Intervention Title | Misconception or fundamental knowledge being tackled | Source | Style | Supporting research |
|---|---|---|---|---|
| Catching a cold | **Catching a cold**<br>Due to the frequency with which this disease is experienced by the public, media and old wives' tales have created misconceptions regarding this particular viral disease. These misconceptions were identified in the research listed and used to inform distractor options. | Self-made, based on research | Two-tier formative assessment probe (MCQ and written explanation) | Johnson and Bungum, 2013: Analysis of common misconceptions associated with catching a cold included getting ill due to the weather or having wet hair. These misconceptions appeared to be robust and held until adulthood without intervention. |
| Food poisoning | **Food Poisoning**<br>This question targets understanding of how bacterial infection can be spread via food, and the infection process by bacteria. | BEST resource, provided by the University of York. | Two-tier formative assessment probe (two MCQs, assessing surface knowledge followed by supporting explanation). | Bandiera, 2007: Misconceptions are strong for the topic of routes to infection, as student knowledge is formed prior to school tuition and can be long lasting<br>Barenholz and Tamir, 1987: Students aged 15-17 were reported to hold animalistic and anthropomorphic views about microorganisms that could 'walk', 'eat' and 'poison us'. |

Lesson 3: The spread of disease

| Intervention Title | Misconception or fundamental knowledge being tackled | Source | Style | Supporting research |
|---|---|---|---|---|
| Managing measles | **Managing measles** Building upon student-led research of the measles virus, research was used to identify common misconceptions associated with viral diseases that then formed distractors of the MCQ. This included the misconception that viruses are alive. | Self-made, based on research | Two-tier formative assessment probe (MCQ and written explanation) | Simon, Enzinger, and Fink, 2017: Students held misconceptions regarding the nature of viruses, naming them the 'evil virus cell'. Students commonly believed that viruses were alive and that they could have bad intentions. |

Lesson 4: Non-communicable diseases

| Intervention Title | Misconception or fundamental knowledge being tackled | Source | Style | Supporting research |
|---|---|---|---|---|
| Can he catch it? | **Can he catch it?** This MCQ hopes to clarify the difference between communicable and non-communicable disease, and in doing so also checks student recall that heart disease is not infectious. The question hopes to address the misconception that all diseases are infectious. | Modified BEST resource | Two-tier formative assessment probe (MCQ and written explanation) | Sigelman, Maddock, Epstein and Carpenter, 1993: Children often base their understanding of all diseases upon common (infectious) childhood diseases, leading to misconceptions<br>Bares and Gelman, 2008: Young children only start to identify that cancer is not communicable, yet colds are at the age of 10. This is of importance as the study group for this research are aged 12-13. |

Lesson 5: Risk factors and data interpretation

| Intervention Title | Misconception or fundamental knowledge being tackled | Source | Style | Supporting research |
|---|---|---|---|---|
| Cause and Effect | **Cause and effect**<br>Linked to the concept of risk factors, this question is rooted in public health research. Misconceptions regarding risk factors include a perceived certainty or immunity to disease due to lifestyle choices. Understanding risk factors has a measurable impact on public health statistics. | Modified BEST resource | Two-tier formative assessment probe (MCQ and written explanation) | Pelikan, Ganahl and Roethlin, 2018: Increased health literacy can cause changes in public behaviour and be beneficial to health outcomes.<br>Kilgour, Matthews, Christian and Shire, 2015: This research recognizes the important role of schools and their curriculum in increasing health literacy<br>Hanson and Gluckman, 2011: Educating children about the development of non-communicable diseases impacts health statistics |

Lesson 6: Genetic Diseases

| Intervention Title | Misconception or fundamental knowledge being tackled | Source | Style | Supporting research |
|---|---|---|---|---|
| Passing it on | **Passing it on**<br>This question provides an example of a child being born with a disorder and suffering from this disease despite living with non-biological parents. Students must identify that the disease is genetic, rather than non-communicable or communicable. This tackles the common misconception that all diseases are communicable. | Modified BEST resource | Two-tier formative assessment probe (2 MCQ questions – one outlining understanding and the next providing explanation) | Piko and Bak, 2006 and Isik, Çetin and Özarslan, 2017: When children aged 8-15 (across these two studies) were questioned about cause of disease, none reported genetic factors.<br>Raman and Gelman, 2005: Student identification of genetic factors causing disease increased with age across a sample of 5-11-year olds, but the permanence of the disorder was often used by students to identify diseases as genetic rather than incorporating a fundamental understanding of genetics. |

Lesson 7: Disease Treatment

| Intervention Title | Misconception or fundamental knowledge being tackled | Source | Style | Supporting research |
|---|---|---|---|---|
| Who needs antibiotics?<br><br><br><br><br><br>Will it work? | **Who needs antibiotics?**<br>Students are presented with different patients, and must decide who needs antibiotics based on their learning. Only one patient has a bacterial infection.<br>**Will it work?**<br>Linked to the previous probe, this tackles misconceptions relating to the action of antibiotics and starts to introduce antibiotic resistance. | Modified BEST resource | Two-tier formative assessment probe (MCQ and written explanation) | Pelikan, Ganahl and Roethlin, 2018: Increased health literacy can cause changes in public behaviour and be beneficial to health outcomes. This is especially important due to the threat antibiotic resistance presents. |

## Appendix One References

Bandiera, M. (2007). *Micro-organisms: everyday knowledge predates and contrasts with school knowledge*. In Pintó, R. & Couso, D. (eds.) Contributions from Science Education Research. Berlin: Springer.

Barenholz, H. & Tamir, P. (1987). *The design, implementation and evaluation of a microbiology course with special reference to misconceptions and concept maps*. In Novak, J. D. (ed.) Proceedings of the 2nd International Seminar: Misconceptions and Educational Strategies in Science and Mathematics, 26-29 July. Ithaca, N.Y.: Cornell University.

Bares, C. B. & Gelman, S. A. (2008). Knowledge of illness during childhood: making distinctions between cancer and colds. *International Journal of Behavioral Development*, *32*(5), 443-450.

Berk, L. E. (2000). *Child Development*. (5th ed.) Boston, M.A.: Allyn & Bacon.

Brindal, E., Hendrie, G., Thompson, K., & Blunden, S. (2012). How do Australian junior primary school children perceive the concepts of "healthy" and "unhealthy"? *Health Education*, *112*(5), 406-420.

Çetin, G., Ozarslan, M., Isik, I., & Eser, H. (2013). Students' views about health concept by drawing and writing technique. *Energy Education Science and Technology Part B: Social and Educational Studies*, *5*(1), 597-606.

Hanson, M. & Gluckman, P. (2011). Developmental origins of noncommunicable disease: population and public health implications. *The American Journal of Clinical Nutrition*, *94*, 1754S-1758S.

Isik, E., Çetin, G. & Özarslan, M. (2017). Students' views about disease concept: drawing and writing technique. *Asia-Pacific Forum on Science Learning and Teaching*, *18*(2).

Johnson, M., & Bungum, T. (2013). Identifying and Reconstructing Common Cold Misconceptions Among Developing K-12 Educators. *American Journal of Health Education*, *44*(3), 169-175.

Kilgour, L., Matthews, N., Christian, P., & Shire, J. (2015). Health literacy in schools: prioritising health and well-being issues through the curriculum. *Sport, Education and Society*, *20*(4), 485-500.

Maxted, M. A. (1984). Pupils' prior beliefs about bacteria and science processes: their interplay in school science laboratory work. Retrieved from: https://open.library.ubc.ca/cIRcle/collections/ubctheses/831/items/1.0096380

Pelikan, J. M., Ganahl, K. & Roethlin, F. (2018). Health literacy as a determinant, mediator and/or moderator of health: empirical models using the European Health Literacy Survey dataset. *Global Health Promotion*, *25*(4), 57-66.

Piko, B. F. & Bak, J. (2006). Children's perceptions of health and illness: images and lay concepts in preadolescence. *Health Education Research, Theory and Practice*, *21*(5), 643-653.

Raman, L. & Gelman, S. A. (2005). Children's understanding of the transmission of genetic disorders and contagious illnesses. *Developmental Psychology*, *41*(1), 171-182.

Reeve, S. & Bell, P. (2009). Children's self-documentation and understanding of the concepts 'healthy' and 'unhealthy'. *International Journal of Science Education*, *31*(14), 1953-1974.

Sigelman, C., Maddock, A., Epstein, J., & Carpenter, W. (1993). Age differences in understandings of disease causality: AIDS, colds, and cancer. *Child Development*, *64*(1), 272-84.

Simon, U.K., Enzinger, S. M., & Fink, A. (2017). "The evil virus cell": Students' knowledge and beliefs about viruses. *PLoS ONE, 12*(3), E0174402.

Wang, S. M., Zou, J.L., Gifford, M., & Dalal, K. (2014). Young students' knowledge and perception of health and fitness: a study in Shanghai, China. *Health Education Journal*, *73*(1), 20-27.

Blank Page