How does the Frayer Model impact concept image formation of Year 9 pupils in Mathematics?

Kathryn Aveyard

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

DEPOSIT AND CONSULTATION OF THESIS

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

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

Surname	Aveyard
First Name	Kathryn
Faculty Board	Education
Title of Dissertation	How does the Frayer Model impact concept image formation of Year 9 pupils in Mathematics?

Declaration by the candidate as author of the dissertation

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

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

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

How does the Frayer Model impact concept image formation of Year 9 pupils in mathematics?

Contents

Introduction

This project was inspired by the conclusions of my Part 2 project. It transpires that there are numerous benefits to both English as Additional Language (EAL) and native English-speaking learners of being explicitly taught the academic language of mathematics. Having previously neglected this element of teaching, it has inspired this project where I trial a recommended approach to vocabulary development.

The Frayer Model, formed in 1969, was designed to measure the conceptual understanding of pupils (Frayer et al. 1969). It is a template with a central oval that should contain a key word, surrounded by quadrants labelled 'My Definition', 'Characteristics/Facts', 'Examples' and 'Non – Examples' (see Figure 1). It has since been recommended to teachers across subject areas in online professional literature, including the National Council of Teachers of Mathematics, as a tool that helps pupils to develop conceptual understanding alongside vocabulary (Dunston & Tyminski 2013; Peters 1975).

Figure 1. The Frayer Model template (Wanjiru & O-Connor 2015, p.203).

The Frayer Model was chosen from a range of possible approaches because it possesses the attributes of an effective language intervention that were identified in my Part 2 study (Author, 2017), whilst having the apparent potential to target another goal I have for my pupils; to develop their conceptual understanding (Adams 2010; Dunston & Tyminski 2013). However, there is no substantial theoretical or empirical evidence to support this latter claim. Theoretical literature exists that, when applied to the model, suggests this could be true in mathematics. However, there is a scarcity of supportive empirical research and what there is bears little relevance to my situation, is outdated, and/or is unrealistic in its approach to teaching (Monroe & Pendergrass 1997; Wanjiru & O-Connor 2015; Heemsoth & Heinze 2014).

As an advocate of research – based practice, I would be more confident in using the model and recommending it to others if there were relevant studies revealing how the model can impact the conceptual understanding of pupils. The results could justify using the model over other learning activities and provide motivation to change my current practice. This project seeks to find such data.

Justification for the focus

Background of part 2 assignment

My part 2 assignment was motivated by a class of Chinese pupils studying A-level mathematics in English. They were unable to comprehend questions independently and therefore could not exhibit their abilities in examinations. I used educational literature to explore the impact that studying in a secondary language has on learning mathematics. The project, largely based on the work by Cummins (2000, 1976), Dawe (1983) and Clarkson (1992), concluded that the level of language proficiency one attains can impact cognitive ability and that developing language can raise attainment in mathematics (Author, 2017). The literature also revealed that the technical language of mathematics is new to native English speakers as well as EAL learners and so explicit language teaching in mathematics can benefit all pupils (Lave & Wenger 1991). Interviews with colleagues were conducted as part of the study. They revealed that my colleagues apply personal theories based on intuition to inform teaching strategies that promote the language development of pupils. In this project, I hope to trial an intervention and collect data to measure its efficacy.

From the outset, the literature was searched for a strategy that resonated with findings from my part 2 project and that could realistically become part of my teaching practice. Therefore, it needed to satisfy some additional criteria. For instance, the approach should be able to be used across many topics in mathematics. For practical and ethical reasons, the approach must be sustainable in terms of resources and preparation time (BERA 2018). Furthermore, it should be easy to implement and to attain a successful outcome so that it can be recommended to colleagues.

The Frayer Model meets these criteria. The template can be used across any domain, extending to any topic within mathematics and according to Adams' (2010) study is also a favourite strategy amongst teachers because it is effective and easy to implement. This study involved 12 teacher-volunteers of secondary mathematics and science working in rural communities in the US. The teachers were taught research-based language strategies and then observed and interviewed to see the uptake of and how they used these strategies. Each teacher tended to consistently use a few of these stratgies, adapting them to align with their views and typical practice. The Frayer Model was a popular choice. Teachers mentioned that the model was effective, well-structured and convenient. It will now be shown, by briefly referring to literature from my part 2 project, that the model satisfies the recommendations for an effective language development intervention (Author, 2017). Much of the literature has an EAL focus, but can still be applied to all pupils (Issa & Eve 2011).

Attributes from part 2 project that are applicable to the Frayer Model

The Frayer Model focuses attention on the key vocabulary to be learnt. The aim of understanding a term is made obvious to the pupil. Mathematics teachers often adopt the view that pupils will `absorb' new vocabulary from the teacher, without the need for explicit teaching (Concepcion 2012). However, researchers believe that making the learning of mathematical language a clear goal for learners is a better approach (Concepcion 2012; Monaghan 2009; Adler 1999); which this model achieves.

The template gives context to the key word by representing it in different forms. For instance, the example may base the word in a real-life context or in the context of an exam question. Whereas the word definition may offer a generalised meaning. Giving a word context makes its meaning more accessible to pupils and therefore easier for them to learn (Concepcion 2012; Monaghan 2009). If the model is completed collaboratively, pupils may also derive meaning by how it is represented by others. They may use, for instance, their hand gestures or other expressions to communicate their meaning of the word (Cummins 2000; Cummins 1976; Cline & Frederickson 1996). This is a different way in which adding context can make the meaning of the word more accessible.

Depending on how the model is used, it can engage pupils in a cognitively challenging task. If pupils are involved in providing the content of the quadrants, it will prompt them to reflect upon and recall their knowledge of the concept. Monaghan (2009) and Cummins (2000) believe that language development should be cognitively challenging and should not be a separate aim to the learning of mathematics. Language is best learnt when there is a purpose for using it; then the language will be heard, read, written and spoken with meaning (Webb & Webb 2008; Barwell 2011; Lave & Wenger 1991).

There is an empirical study that supports the model's ability to promote language development. Monroe and Pendergrass (1997) use an adapted version of the Frayer Model to teach about measurements in their study. As seen in Figure 2, this version contains the quadrants of the Frayer Model in a slightly different format. The vocabulary to be targeted during the intervention was selected by a team of teachers. The Frayer Model was used in a variety of ways to reflect the reality of teaching; sometimes it was used to stimulate class discussion and sometimes for individual reflection of a concept. The results showed with significance that pupils using the Frayer Model could list more concepts with measurement content than those pupils only offered the definition during lessons. This suggests that the Frayer Model is a more effective method of developing language than use of definitions alone.

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at Monroe, E. & Pendergrass, M., 1997. Effects of Mathematical Vocabulary Instruction on Fourth Grade Students

Figure 2. Illustration of the adapted Frayer Model used in Monroe and Pendergrass' (1997, p. 24) study

It has now been shown that theoretical and empirical support exists for the model's ability to promote language acquisition. Some believe that language influences how information is processed and is therefore inextricably linked to the development of conceptual understanding (Adams 2010; Cummins 2000; Vygotsky 1986). It may therefore seem unsurprising that a model designed to support language acquisition can also develop conceptual understanding. This project aims to look at how the model may promote the development of conceptual understanding independently of this language focus. To begin, the term 'conceptual understanding' will be defined with respect to its attributes This will highlight why it is a learning goal I have for my pupils and why it is the focus of this study.

Meaning of conceptual understanding

Conceptual understanding is a type of mathematical knowledge that is required to become proficient in mathematics (Rittle-Johnson & Alibali 1999; Kilpatrick 2001). This category of knowledge was identified to help clarify the goals of learning in mathematics and thus what we aim to assess (Rittle-Johnson & Alibali 1999). There are many definitions for this term across the literature and they seem to draw on three key indicators or attributes which will now be described.

Firstly, to have conceptual knowledge is to have connected knowledge (Star 2005). Many describe a concept as a network or web of connected nodules of isolated information (Klausmeier 1992; Jennings 2012; Freeman et al. 2003; Evrekli et al. 2010). The nodules could comprise of definitions, procedures, examples or experiences of the concept that are internal or external to school. The notion of a web acknowledges that one must be aware of the relationships between them; to recognise their similarities, differences and how changes in one may impact the other. It includes having a knowledge of procedures as well as how they work and how to adapt them to different situations (Skemp 1976; Rittle-Johnson & Alibali 1999). Many connections allow many ways to recall knowledge and so there is a reduced dependency on memorising isolated nodules of information (Skemp 1976; Coutinho 2014).

Conceptual understanding is also seen as dynamic. Understanding is not a state that can be achieved (Watson 2006; Klausmeier 1992). It grows and is enriched as new knowledge and experiences are encountered (Tall & Vinner 1981; Klausmeier 1992; Warren & Cooper 2009; Coutinho 2014; Carpenter & Lehrer 1999). The way new knowledge is integrated depends on which theory of learning one subscribes to. As conceptual understanding cannot be attained, Klausmeier (1992) has created hierarchical levels that can indicate the depth of understanding achieved: concrete, identity, classificatory and formal (see Appendix 1).

Finally, conceptual understanding suggests awareness of the limits and principles defining a concept (Rittle-Johnson & Alibali 1999; Lefevre & Hiebert 1986; Mason & Watson 2005). The limits that one can identify will depend on their level of experience with that and any connected concepts. For instance, a pupil may not specify that a square is a 2D shape if they do not have an explicit awareness of 3D shapes. Furthermore, the definition of square may vary depending on the context; it can be defined in terms of vectors, lengths, angles, symmetry etc. Awareness of the critical features of a concept is associated with Klausmeier's (1992) higher levels of conceptual understanding and is desirable because it is decontextualized and thus more transferable to unfamiliar problems (Carpenter & Lehrer 1999; Lefevre & Hiebert 1986; Klausmeier & Frayer 1970).

By defining conceptual understanding in the context of mathematics, some of its attributes have now been identified; it allows better recall of knowledge and better prepares pupils to apply their knowledge to unfamiliar situations (Skemp 1976; Lefevre & Hiebert 1986). These attributes can help to raise the attainment of pupils when assessed in terms other than their ability to perform basic procedures. Therefore, the development of conceptual understanding in mathematics is the focus of many interventions and recent policies (Cockcroft 1982; Bisson et al. 2016). The professional literature presents the Frayer Model as a convenient blueprint that teachers can use to promote this (Dunston & Tyminski 2013). However, this has little gravitas unless it is supported by empirical evidence showing that it can be effective in a typical classroom. An exploration of how the model impacts concept formation should provide further insight and perhaps more confidence in the model.

Aims and design of this project

To assess the impact of using the Frayer Model as a learning tool to develop a pupil's conceptual understanding, a valid and reliable way of observing this must be sought. Thus, there are two key foci in this project which will now be stated alongside the approach to addressing them:

1) To find a way of observing the development of conceptual understanding

There are two elements of this aim. Firstly, a valid assessment task that allows pupils to exhibit their understanding of a concept. Secondly, a reliable and valid way of evaluating the assessment task. This study uses mind maps as an assessment task and these are evaluated quantitively and qualitatively to answer the following three sub-questions:

1) How effective is the Frayer Model in developing conceptual understanding?

The purpose of this question it to find evidence that would justify and motivate a change to my current teaching practice.

2) Is there evidence of the impact of the model in pupils' mind maps?

This is to see how the Frayer Model impacts pupils' development of conceptual understanding.

3) What are the qualities of the students' mind maps that made the most and least progress in their conceptual understanding?

The purpose of this question is to observe the range of responses pupils have to the intervention.

2) **To observe how the Frayer Model impacts a pupil's understanding of a concept**

Ultimately the title question will be answered using practitioner research conducted by me in my own school. Prior to this, a review of relevant empirical and theoretical literature will be conducted. The literature will be reviewed with regards to each quadrant because there is a lack of research focused specifically on the Frayer Model. The aims of this review will be to seek guidance regarding implementation of the model and to gain insight as to why the model is believed to support conceptual understanding. These are the aims because it is risky to change normal practice without understanding what is required during implementation to promote the development of conceptual understanding.

Research-informed standpoint on project's foci:

1) To find a way of observing the development of conceptual understanding

To observe the change in conceptual understanding, developmental data must be collected that shows how the level of a pupil's conceptual understanding has changed. This may be achieved by observing how their concept image changes over time. Vinner (2002) defines 'concept image' as

‘…the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes. It is built up over the years through experiences of all kinds, changing as the individual meets new stimuli and matures.’(Vinner 1991, p.152)

The concept image is like a snapshot of one’s current conceptual understanding. It comprises all their knowledge, including procedural and conceptual, that they associate with the name of the concept. This definition is significant because it suggests that changes in pupils’ conceptual understanding can be observed if there is a way to assess their concept image before and after the intervention.

This section will be structured in two parts. The first will review recommended assessment tasks and justify the use of mind maps in this project. The second will regard methods of evaluating mind maps. The task and evaluation approach will be reviewed with respect to their validity in terms of being able to capture one’s concept image and their associated workload.

An online search suggests that there are currently no other studies regarding how the Frayer Model impacts conceptual understanding in Mathematics. If there were, then it may have been useful to mimic their assessment approach so that results could be compared. There exists one relevant study conducted in Kenya by Wanjiru & O-Connor (2015). They explore how the model can impact achievement in Mathematics. Since measuring achievement can sometimes align with measuring conceptual understanding, the assessment task used in this study is of interest and will also be considered.

Assessment task

The aim of Wanjiru & O-Connor's (2015) study is similar to this one; their aim was to assess the impact of vocabulary instruction on achievement. They chose the Frayer Model because of its potential to engage pupils and develop understanding alongside vocabulary acquisition. They involved two secondary schools - one boys and one girls - in the same location, with similar facilities, and similar results in standardised tests. They used random stratified sampling to maintain equal numbers of girls and boys in their control or experimental groups. Wanjiru & O-Connor (2015) used a total of five assessment tasks. Pre- and post- mathematics achievement tests, pre- and post-mathematics vocabulary tests and dictionaries that pupils created throughout the process. The achievement tests comprised of adapted Kenyan national exam questions that used the vocabulary taught in the intervention. There were five questions in the pre-test made to coincide with the five levels of Bloom’s Taxonomy. There were ten in the post-test, two per the five levels of Bloom’s Taxonomy.

Although national tests may be a good indicator of achievement, they are arguably not the best way of measuring conceptual understanding. Some feel that the validity of these tests is compromised to ensure that they can be marked reliably by examiners with different background across the country (Jones & Inglis 2015; Price & van Jaarsveld 2017). They suggest that the questions are designed to elicit a small number of predictable responses that can consistently be marked according to a mark scheme. McClure et al. (1991) believe these questions limit a pupil’s ability to express knowledge and understanding. Perhaps Wanjiru & O-Connor's (2015) careful selection of exam questions that coordinate with Bloom’s Taxonomy mean that their particular assessment is a good measure of a pupil's understanding. But there are other, more open-ended tasks that appear better able to capture a pupil's concept image.

Tests involving open-ended questions are often used as an option as these result in a wider variety of responses. Pupils can solve them in many ways and the deductions they make provide insight about their understanding of relationships between concepts (Bisson et al. 2016). However, there are two drawbacks of using open-ended questions. These questions need to be carefully designed and tested to ensure that the responses do reveal a pupil's understanding (Riccomini et al. 2015; Price & van Jaarsveld 2017). This requires great skill to do; Bisson et al. (2015) suggest that there are tests with proven validity to measure conceptual understanding, but only for a small number of topics. The second issue they add is that the wide array of responses is difficult to evaluate consistently. This is because it is much harder to predict the answers and thus make a mark scheme. The study's validity will be limited if responses cannot be marked reliably.

Portfolios and journals are also popular open-ended tasks that have some similar advantages and disadvantages to using open-ended questions. An additional advantage they have over pre- and post- testing is that they are completed over a period of time and can therefore show nuanced shifts in understanding (Smith 1993). Furthermore, they give pupils freedom in their method of communication (Smith 1993); they can write, draw diagrams, create examples etc. which is useful for topics that are best represented pictorially, like Circle Theorems. There are some criticisms of journals/portfolios aside from the slightly greater burden on the teacher of having to manage storing them over the assessment period. McClure et al. (1991) argues that creating journals/portfolios is itself a skill and that a pupil's performance will therefore be influenced by their conceptual understanding and their ability to demonstrate their knowledge in this format. Further issues exist when evaluating these tasks. Journals/portfolios provide are a large and rich data source; effectively evaluating a large sample size can be time consuming (Short 1993). Creating a marking rubric would increase efficiency and reliability but may be difficult to create and an overly-simplistic way of evaluating a complex construct like conceptual understanding (Bisson et al. 2016).

Skemp (1976) believes that interviews are the best way of assessing conceptual understanding. This is because a skilled interviewer can ask follow-up questions to ensure the question has been understood and to deeply probe an individual's understanding (Powney & Watts 1987). Interviews have some of the disadvantages already mentioned. For instance, they provide a wealth of data that is time-consuming and difficult to fairly evaluate (Powney & Watts 1987). Furthermore, the response provided is dependent on the pupil's skill with articulating their thoughts in that moment (McClure et al. 1991). There are additional disadvantages relevant to this project. To be effective at probing understanding, the interviewer must ask appropriate and unbiased follow-up questions. Formulating these questions whilst carefully listening to a response is challenging and made more difficult as a subjective practitioner researcher (Bisson et al. 2016; Reeves et al. 2010). Although interviews have merit in assessing conceptual understanding, they are not suitable for all circumstances.

The Frayer Model itself was first introduced as a way of assessing conceptual understanding (Frayer et al. 1969). Many of the advantages and drawbacks of other assessment tasks already mentioned are applicable to the model. However, the main reason that this is not the method of choice is because it is also the intervention task. Pupils in the intervention group would be advantaged. Having experienced it before, they may be more advanced in the skill of the task. Furthermore, it will be easier for pupils to simply memorise and regurgitate the models created in class without necessarily having understanding; an issue raised by Vinner (1991). Although the model offers a valid assessment task, it would not be suitable to use for this project.

Mind maps will be used to assess conceptual understanding in this project. A 'mind' or 'spider' map is a type of concept map that was attributed to Joseph Novak in 1972 (See Figure 3) (Turkay n.d.; Coutinho 2014). A mind map has a concept in the centre and connecting knowledge attached by lines. More sophisticated maps describe the relationship between the pieces of knowledge along the connecting lines, reflecting the structure of a concept image.

Figure 3. Different forms of concept mapping (Turkay n.d., p.6)

Mind maps were chosen because they are a valid way of assessing conceptual understanding and require minimal skill and time to do. Like journals/portfolios, pupils can use a variety of mediums to communicate their understanding (Abi-El-Mona & Adb-El-Khalick 2008; Price & van Jaarsveld 2017). The pupil can offer examples, drawings, definitions and are not limited by the context of the task (McClure et al. 1991). The structure of it can also be revealing; the connections can show the deductions the pupil has made and so misconceptions or knowledge that has not been integrated can be easily identified (Freeman et al. 2003; McClure et al. 1991). Unlike the other assessment tasks, creating mind maps is a relatively simple skill taking very little time for pupils and little time, skill and organisation on behalf of the researcher to prepare and conduct (McClure et al. 1991).

McClure et al.'s (1991) study offers evidence supporting the validity of mind maps. They use various methods to evaluate and score a sample of mind maps. They compared these scores against a master mind map created by experts and concluded that the scores given were reflective of their level. Since the master mind map is subjective and based on the knowledge and views of the experts, it would still be of interest to see how the results from mind mapping compare with other already validated methods.

Aside from validity, mind maps are also an ethical choice. They are often used solely in classrooms as a learning task. They are believed to help pupils to reflect and organise their thoughts creating clarity and interconnected thinking (Freeman et al. 2003; Novak

1990; Turkay n.d.). Pupils may therefore benefit from participating in this assessment process.

There are some drawbacks to using mind maps. Although mind maps are shorter than portfolios/journals, they are a rich data source that can be time-consuming and difficult to reliably evaluate (Evrekli et al. 2010). Furthermore, although they are praised for allowing pupils to show the relationships they have made, Coutinho (2014) suggests the design does not allow more complex connections to be noted. It seems possible that this could be overcome by allowing more, complex relationships to be represented as a separate nodule.

To summarise, the assessment task of the most relevant empirical study that uses the Frayer Model has been considered as well as others that are believed to provide insight into a pupil's concept image. Mind maps have been selected because they appear pragmatic and valid. However, there are concerns about how these may be reliability evaluated to produce valid results. Literature will now be reviewed to discover how mind maps have been evaluated in previous studies, before justifying the use of the comparative judgement method.

Evaluation of the assessment task

Decisions around the evaluation method depend on the sample size and what is being measured. Qualitative evaluations better account for detail in rich assessment tasks as they can reflect the complexity of measuring conceptual understanding (Bisson et al. 2016). However, this is too time consuming with a large data set and offers too much analysis to process (Edwards & Talbot 1994). Extracting quantitative data can be viewed as oversimplifying the assessment process, but it provides a useful overview of results and can direct the researcher to check certain responses (Jones & Inglis 2015). This study will evaluate the mind maps quantitively and qualitatively and thus the literature will now be searched for an appropriate way of attaining quantitative data from mind maps.

Marking rubrics have been used by many to score mind maps. These rubrics are usually based on Lefevre and Hiebert's (1986) belief that depth of understanding is correlated to the number of connections on a mind map. Figure 4 presents an example of such a scoring system; Evrekli, Inel and Balim (2010) offer many more in their paper. However, these rubrics do not always reflect the quality of the relationships described. A pupil that makes many simpler points may attain a higher mark than another that identifies fewer but more complex relationships (Freeman et al. 2003). With its validity in question, Coutinho (2014) believes more convincing evidence showing the validity of marking mind maps with rubrics needs to be found.

Figure 4. An example of a scoring system designed to assess mind maps (Evrekli et al. 2010, p.2331)

McClure et al. (1991) analyses the appropriateness of holistic and relational approaches to evaluating mind maps. These methods supposedly account for the complexity of rating the content and structure of mind maps. Holistic evaluation is when the whole map is judged based on the marker's expertise and allocated a score from 1 – 10. The criteria for relational marking is harder to describe and can be seen in Figure 5. McClure et al. (1991) tested the validity, reliability and marking time required of these approaches.

The holistic approach appears to be superior to the relational in terms of validity, reliability and time efficiency. McClure et al. (1991) tested the validity of these methods by comparing the scores with a master mind map, created by experts. The holistic approach had a Pearson Product Moment Correlation Coefficient (PPMCC) of .387, with the relational having a PPMCC of r=.379. Both indicate that the marks produced by following these evaluation systems are not consistent with the expectations of the experts that made the master mind map. This implies that these evaluation methods have a low validity rating. However, it could be argued that the PPMCC only measures how much the pupils' work reflects the particular master mind map used for this study. A different group of experts may have produced a different master mind map that may have led to a higher validity rating. When the markers had access to the master mind map whilst scoring, predictably McClure et al. (1991) reported higher PPMCC ratings. Reliability was tested using the GENOVA program; the holistic approach (g=.67) was deemed more reliable than the relational approach (g=.51). Finally, the holistic approach was also much faster, taking only 85 minutes to mark 63 mind maps, compared to 169 minutes using the relational method. Therefore, the holistic approach is superior to the relational approach in terms of validity, reliability and efficiency.

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at McClure, J., Sonak, B. & Suen, H., 1991. Concept Map Assessment of Classroom Learning: Reliability, Validity, and Logistical Practicality. Journal of Research in Science Teaching1, 36(4), pp.475–492

Figure 5. Criteria for relational marking (McClure et al. 1991, p.482)

Due to Klausmeier's (1992) paper, where he created levels to clarify different depths of conceptual understanding, it was considered whether this project should aim to assess pupils' level of understanding rather than allocating a numerical score. The level descriptors can be seen in Appendix 1. Levels are more holistic and seem a better indicator than a number score in representing the complexity of conceptual understanding. Progress could be indicated by pupils moving up the levels. However, they were not used in this project for the following reasons. The descriptors are generic to any topic in mathematics and would need to be interpreted in the context of the topics taught in this intervention. This would make the process of marking much more complex and cognitively challenging, which is likely to be quite time consuming. In addition, progress would be difficult to show because there are only four levels. With only four, it is harder to differentiate between the progress made by pupils.

The evaluation method used in this project is called comparative judgement (CJ) and is relatively new to the Education community. It is another holistic method based on the expertise of markers. By repeatedly comparing pairs of mind maps, or scripts, and deciding which is best, the scripts are allocated a score. There is a website that supports this process, accessible for free at www.nomoremarking.com. The website, created by educational researchers, presents pairs of scripts according to an algorithm. The algorithm minimises the number of comparisons required to reliably score the scripts (Jones & Inglis 2015; Jones et al. 2014). Figure 6 shows a screenshot example of what a judge would be presented with. Multiple judges can participate in the marking process. Each judge can be allocated a quota of comparisons, reducing individual workload.

The CJ website generates a lot of data and not all of it is useful for this particular project. This study will focus on the scores generated and the reliability or consistency of marking. The following information has been ascertained from the website's blog and from emails with the creators of the website (NMM 2018). The 'true score' is generated by an algorithm and is based on the judgements made by the markers. The true score is manipulated to always have a mean of zero and values that can extend from negative to plus infinity. The scaled score reported is a linear translation of the true score, ranging from 0-100. The website also generates an overall reliability rating result using the Rasch Reliability measure. The reliability is based on the consistency of marking across difference judges. The use of the website can make the marking of mind maps an efficient, reliable and collaborative process.

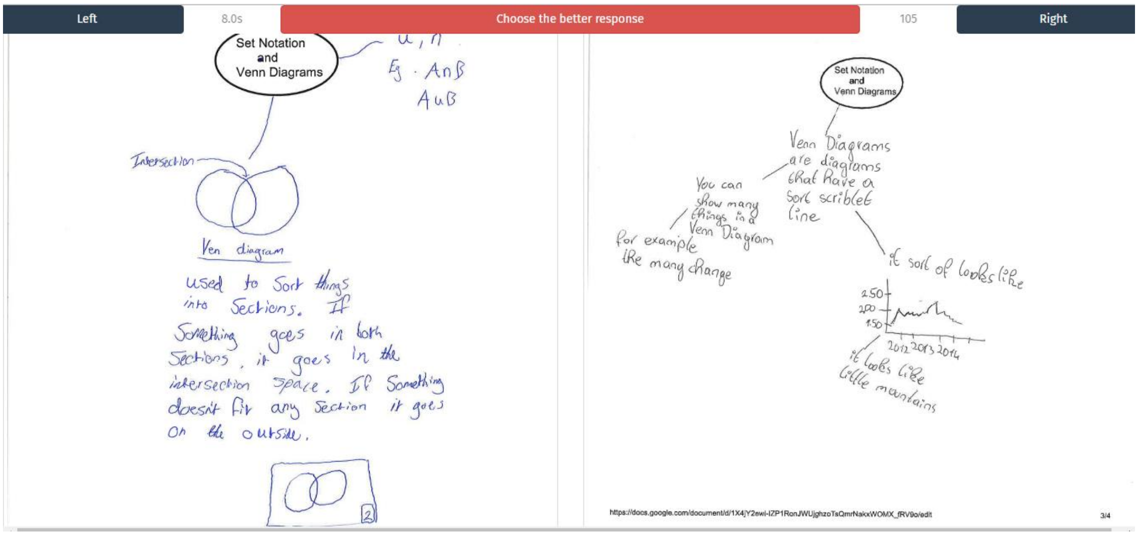


Figure 6. Screenshot from nomoremarking.com showing how work is presented. Experts pick their preference, 'Left' or 'Right'.

Comparative judgement has mainly been chosen for this project because of its good validity rating. This holistic approach benefits from using the knowledge of experts to identify markers of understanding that a rubric might miss (Smith 1993; Jones & Inglis 2015); which is particularly relevant to mind maps where both the structure and contents must be accounted for (McClure et al. 1991). It also more reliable because it is less dependent on having marking experience; it is much easier to compare two scripts and decide which is better than to allocate a score or attempt to apply a mark scheme (Jones et al. 2014, Thurstone 1994 as cited in Bisson et al. 2015). In addition, a benefit of choosing to mark collaboratively is the decreased likelihood of bias caused by a single judge favouring certain content or methods of presentation (Bisson et al. 2016).

There is statistical support of the method's validity. Jones & Inglis (2015) challenge the validity of national examinations and present an assessment approach they believe is more suitable to measuring understanding in mathematics. They employed GSCE question writers to create GCSE problems that do not need to be restricted by the need for a mark scheme. Teachers confirmed that the paper they created tests problem solving skills and content knowledge. A total of 750 pupils from UK secondary schools sat the paper. Markers with a mathematical background were recruited and paid to participate. Comparisons between the judgements made by different markers indicated a high level of internal consistency. The validity of their results was checked using the PPMCC with pupils' predicted GCSE marks ($r=0.86$) and grades ($r=0.71$). Thus showing that the results of marking questions with CJ is comparable to marking with a mark scheme, but CJ has the added benefit of greater freedom in the task design. Similar conclusions were made by Bisson et al. (2015) who compared their results with other valid measures of attainment. They too found that CJ correlated with other validated forms of assessment.

The comparative judgment method has some limitations. Earlier it was argued that having multiple markers reduces the tendency for bias. However, it is still susceptible to systematic bias - where all the markers have the same preferences (Jones & Inglis 2015). This risk may be minimised by recruiting judges with diverse backgrounds and principles. One drawback of the website is that it only displays one A4 side of pupils' work, giving a narrow overview compared to a whole national examination (Jones et al. 2014). Since the results generated by the website have now been shown to align with predicted scores and grades on nationalised tests, this concern seems unfounded (Jones & Inglis 2015). In any case, this is not applicable to this project where the intervention will only be trialled for one topic.

To summarise, the aim of this section was to find a way to assess the impact of the intervention on pupil's conceptual understanding. It was decided that this would be achieved by trying to capture and analyse their concept image before and after the intervention. Mind maps will be used in this project because they present a valid way of achieving this and are easy to implement. These mind maps will be evaluated in various ways to try and answer the three sub-questions identified earlier. The use of comparative judgement will be one of these ways as the quantitative data attained will give an overview of the progress made by pupils. The marking will be a collaborative process, completed with colleagues. It will then be used to direct a more in-depth qualitative analysis of some scripts.

2) To observe how the Frayer Model impacts a pupil's understanding of a concept

Literature will be used to gain insight into why and in what way the Frayer Model might impact a pupil's concept image. A biproduct might be some guidance about how to use it in the intervention. There is little, if any, literature explaining how the model contributes to conceptual understanding, but there is a plethora of information regarding how the individual quadrants promote it. To frame the rest of the literature, a brief description will be given of how conceptual understanding is formed. Afterwards the literature will be organised according to the quadrants in the model.

Development of the concept image

The development of conceptual understanding is often described in terms of the web or network of knowledge described earlier. Warren and Cooper (2009) suggest that concepts develop when more connections are made between the knowledge nodules. Piaget (1952) adds that these are created when the concept is engaged with or operated on in some way. This is because it requires one to make a judgement about that knowledge, so it can be organised and integrated into the network. Hence, many interventions that aim to develop conceptual understanding aim to raise awareness of relational knowledge. For instance, Klausmeier and Frayer (1969) suggest that creating classifications for objects encourages thinking about their similarities and construction. Carpenter and Lehrer (1999) recommend reflection-based activities, as they encourage recall and reorganisation of knowledge. They also advise articulating, applying and making knowledge one's own. Therefore, if the Frayer Model develops one's concept image, it must require pupils to operate on the concept.

Once information is judged, it must be integrated into the network. Piaget (1952) describes two ways this can happen; assimilation or accommodation. If the new knowledge is coherent with the current concept structure, the new knowledge is assimilated into it, reinforcing it. If the new knowledge is conflicting and cannot be integrated, the structure of the concept must adapt to accommodate it. When either assimilation or accommodation occurs, the whole knowledge structure is impacted (Klausmeier & Frayer 1970). The resultant relationships created from this process reveal further information about the discrete knowledge pieces and therefore create greater understanding in themselves (Piaget 1952).

Klausmeier's (1992) levels indicating depth of understanding align with this process. He correlates depth of understanding with one's ability to abstract the critical and non-critical features of a concept. This can only be done when a sufficient number of connections between relative topics have been made (Klausmeier & Frayer 1970; Vinner 1991; Bills et al. 2006; Bills & Rowland 1999). It may also be the basis for the belief that counting the number of connections on a mind map indicates one's depth of understanding (Lefevre & Hiebert 1986).

Definitions and characteristics/facts

Mathematics is based upon concepts, and concept definitions are society's attempt to translate these into words (Klausmeier 1992, p.269). Consequently, they are believed by some to have significance in the study of mathematics. When discussing a pupil's perception of a concept, it is useful to identify the subcategories of formal definition and personal definition. Formal definitions are agreed upon by experts and tend to identify the critical features of the concept as well as their function and role relative to other concepts (Zaslavsky & Shir 2005; Klausmeier 1992). A personal definition is an individual's version of the formal definition, based on their current concept image (Vinner 1991; Tall & Vinner 1981). How closely the formal and personal definitions align indicates the accuracy of their understanding (Price & van Jaarsveld 2017). It may therefore be possible to support the creation of an accurate concept image by helping pupils to create a correct personal definition.

Offering formal definitions to pupils is not always an effective way of influencing their personal definition. Concepcion (2012), Vinner (2002) and Klausmeier (1992) believe that formal definitions can be challenging for pupils to access as they present concepts in a decontextualized form. Examples of a concept, for instance, are easier to comprehend. Vinner (2002) warns that they may appear useful if a pupil can recite a formal definition, but he has found that some memorise the formal definitions without understanding or using them during problem solving. Their value for usage in lessons is redeemed only slightly by Klausmeier (1992) who believes they can be useful for high ability pupils when studying abstract concepts as they can provide clarity and prevent misconceptions.

Tasks where pupils create their own definitions post-study of a concept are a more popular way of integrating definitions into lessons. Zandieh & Rasmussen's (2010) study presents vignettes of pupil conversations when tasked with defining triangles. Initially the pupils rashly created a definition but with more time they then began to negotiate and refine it, aiming to make it as accurate and succinct as possible. It resulted in discussions about what was a necessary feature of the definition and what was a consequential characteristic/fact, such as the interior angles totalling 180°. Zandieh & Rasmussen (2010) consequently conclude that when pupils create their own definitions, it highlights the critical features of a concept as well as its similarities and differences to neighbouring concepts. Mariotti and Fischbein (1997) and Zaslavsky and Shir (2005) also support this claim, adding that the personal definitions these pupil’s create are more likely to be used when problem solving than the formal definitions they might be presented with.

To summarise, this section has highlighted why emphasising the definition and characteristics/facts of a concept might be useful in lessons. Definitions can add clarity about the necessary features of a concept whilst characteristics/facts list additional properties; which were earlier recognised as attributes of conceptual understanding (Mason & Watson 2005; Jennings 2012). It has also been noted that the understanding of pupil's benefits most by engaging with this process of creating definitions and identifying facts/characteristics. This resonates with Piaget's (1952) belief that conceptual understanding is created when one operates on a concept.

Examples and Non-Examples

Examples can be a more accessible way of representing a concept (Klausmeier 1992). Watson and Mason (2002) define examples as:

'...any situation in which something specific is being offered to represent a general class with which the student is expected to become familiar'. (Watson & Mason 2002, p.239)

For examples to be an effective way of clarifying a concept, the critical and non-critical features must be identifiable. Otherwise, it is possible for pupils to incorrectly abstract the critical features of a concept from an example, resulting in an incorrect concept image. The critical features can be highlighted by either presenting a sequence of examples that demonstrate the features that can and cannot be changed, or by the teacher stating them (Skemp 1976; Bills et al. 2006; Warren & Cooper 2009).

As with definitions, allowing pupils to generate their own examples, instead of always being shown them, can offer additional benefits. When teachers create examples, they sometimes engineer them to achieve a particular focus (Watson & Mason 2002). They can achieve this by manipulating the non-critical features. For instance, when beginning to teach the procedure of factorising quadratic expressions, the teacher may only present examples where the x^2 coefficient is 1 and the other coefficients are simple integers that will allow the quadratic to be factorised easily. The pupil may then incorrectly believe that all quadratics where the x^2 coefficient is 1 can be factorised. Watson & Mason (2002) and Bills et al. (2006) believe that pupils benefit from creating their own examples by becoming aware of these manipulations. They gain greater insight into the structure of the concept, recognising its critical and non-critical features; an indicator of deep understanding (Rittle-Johnson & Alibali 1999; Lefevre & Hiebert 1986; Mason & Watson 2005).

Unlike examples, the use of non-examples in lessons is a more contentious issue. Non-examples are supposed to be almost-examples of a concept (Bills et al. 2006). That is, a non-example is similar to an example but due to the absence or difference in some feature, it is not an instance of the concept. Bills et al. (2006) advocates their use. They believe that clarifying why a non-example is not an example highlights the boundaries of a concept. However, empirical studies have produced mixed conclusions.

A recent study suggests non-examples may not benefit all pupils. Heemsoth and Heinze's (2014) study aimed to discern whether the development of negative knowledge results in a deeper understanding of a concept; where negative knowledge is knowledge of incorrect strategies and concepts (Prediger 2008, cited in Heemsoth & Heinze 2014). It took place in Germany with 217 pupils from four secondary schools. The intervention focused on teaching how to multiply and divide fractions over a three-week period. They used a non-random quasi-experimental design where both the control and experimental group both received the same instruction. However, in the seven practice sessions, the experimental group reflected on correct or incorrect examples as opposed to the control group that only experienced correct examples. The pupils were tested before and after teaching on their knowledge and negative knowledge of fractions.

Pupils in the experimental group developed their understanding of negative knowledge more than the control group. However, this improved the understanding of fractions for only the higher attaining pupils. Heemsoth & Heinze (2014) summarise that pupils with low prior knowledge would gain more benefit from seeing examples only and that non-examples should just be used with pupils that already have an established understanding of the concept otherwise they risk becoming confused and incorrectly integrating the non-example. It is perhaps significant to note that their study focused on examples and non-examples of procedures as opposed to the broader category of concepts that will be used in this study.

Several other studies have echoed the sentiment that non-examples are less effective at developing understanding than examples (Cohen & Carpenter 1980). However, there are criticisms of the methods used in these studies. For instance, MacHale (1980, cited by Bills et al. 2006) believes that non-examples are useful in classrooms where examples are not mindlessly copied but are instead the subject of discussion. Discussions where the features of the example are highlighted. In this type of classroom, a pupil is not likely to mindlessly mimic a non-example. It is possible that the lower attaining pupils in Heemsoth & Heinze's (2014) study were in a class that typically copy and follow procedural examples and this may explain their negative results. It also seems that pupils in the experimental group reviewed the non-examples with their peers as opposed to with further explanation from the teacher which some might argue is not representative of normal lessons; this is an argument made by Cohen & Carpenter (1980) about most of the studies that do not support the use of non-examples.

This section aimed to review why examples and non-examples may support the development of a pupil's concept image and to seek guidance about how best to implement the Frayer Model. It seems that the use of examples can highlight the critical features of a concept and that non-examples develop negative knowledge, which can contribute overall to a more comprehensive concept image (Heemsoth & Heinze 2014; Bills et al. 2006). As with definitions, pupil-generated examples allow pupils to operate on the concept, giving additional benefits and thus this intervention will aim to provide pupils the opportunity to do this (Piaget 1952).

Studies using the Frayer Model

It was noted earlier that the most relevant study using the Frayer Model was conducted by Wanjiru & O-Connor (2015) in Kenya. They looked at its impact on achievement in secondary school mathematics. They used a quasi-experimental design with 216 pupils. The experimental group used the Frayer Model to learn new vocabulary whilst the control group were only given definitions. They found that the girls and boys in the experimental group outperformed those in the control group. They used an independent t-test in their quasi-experimental study. It showed the results were statistically significant $(\rho = .001)$ but the effect size was small $(r = .1875)$. This conclusion is important because at a minimum it suggests that the model is not detrimental pupils. However, as with any study in education, its limitations and relevance to one's own situation must be considered.

Coutinho (2014) and Skemp (1976) suggest that a big determinant of conceptual understanding is a pupil's willingness to seek out connections and relationships when learning. Whilst Klausmeier (1992) lists IQ scores, self-esteem, achievement motivation, self-directedness and attitudes towards school as all having a significant impact. It is important to recognise that the efficacy of the model will be impacted by these factors. It is therefore possible that the small calculated effect size in Wanjiru & O-Connor's (2015) study may be a consequence of the culture of the school's they were working in.

Wanjiru & O-Connor (2015), as well as more recent studies aiming to measure conceptual understanding, do not try to eliminate the influence of these factors from their results (Price & van Jaarsveld 2017; Fuchs et al. 2017). It is not clear why such attempts have not been made. It could be that the process is complex; maybe they believe the impact is negligible or their interventions should be effective regardless of these factors. The purpose of this section is to identify major influences on the development of conceptual understanding and to highlight the need to conduct research at my school, since the results will be specific to my pupils and school environment.

Summary

The purpose of this section was to search literature so that it could inform the assessment process and could provide insight as to how the Frayer Model may impact the development of conceptual understanding. Whilst both these objectives have now been met to some degree, there is a scarcity of literature regarding the model. That is, although the literature has been researched with respect to each quadrant, it is still unknown as to whether engaging with all the quadrants in one activity may have an impact and there are a lot of questions surrounding this; for instance, would pupils refer to their definition and characteristics/facts quadrants when creating their examples? In what order would pupils fill out the model? It is not the purpose of this study to answer these questions, but it is useful to identify them as areas of research that require further exploration.

Practitioner Research

Practitioner research in schools occurs when someone identifies an issue and conducts a study within their own school (Feldman et al. 2007; Babione & Anderkin 2015; Pelton 2010). This study can be described as practitioner research because initially a language barrier was identified as an issue and now this project aims to assess whether a language strategy can be used to develop the conceptual understanding of pupils in mathematics in my school.

Action research is a type of practitioner research that follows a cyclical process of continual improvement and reflection outlined by Wilson (2017) as Plan, Act, Observe, Reflect. This structure was used in this study because it allows the method to be refined.

Three cycles were conducted altogether, with the first being just a pilot of the CJ website. A time line of events can be seen in Appendix 2.

Certain procedures should be followed when conducting research in one's place of work because of the professional relationships they have had and must continue to have there (Babione & Anderkin 2015). This project is therefore informed by literature about research ethics in the social sciences and conducting practitioner research in schools.

Evaluation Criteria

The aim of this research is to observe how the Frayer Model impacts pupils' understanding of a concept. To try and observe these changes, the mind maps will be evaluated with respect to three questions. The purpose is to triangulate the evaluation approach, providing a more thorough picture of how the model impacts the concept image. How the data will be used to answer the three research questions will not be addressed:

1) How effective is the Frayer Model in developing conceptual understanding?

This will be addressed using a pre-/post- test quasi-experimental design. The scores generated from the website as a direct result of the comparison judgements will be used to calculate the average progress in conceptual understanding made by the experimental and control group. The significance level and effect size will also be computed to give an indication of the reliability of the results and how significant the model's influence on concept development is.

2) Is there evidence of the impact of the model in pupils' mind maps?

To observe this, the content of the mind maps will be searched for the number of scripts that contain instances of examples, non-examples, definitions and facts/characteristics before and after the intervention. This will indicate the uptake of particular features of the model which is likely to show a direct impact of having engaged with the model.

3) What are the qualities of the students' mind maps that made the most and least progress in their conceptual understanding?

Due to time constraints and the sample size, it is not possible to conduct an in-depth qualitative analysis of the data. The numerical data generated from the comparative judgement approach will be used to direct this detailed analysis to the pupils that made the most and least progress. The before and after mind maps of these pupils will be compared and analysed to look for developments in their understanding.

Cycle 1 (Pilot)

Pilot studies are used to increase the likelihood that the main study will be successful in achieving its aims. Due to time restrictions, the whole intervention process could not be trialled. The critical part of the process, the use of the www.nomoremaking.com website with mind maps, was the only part tested in this cycle. It was critical to test this because it has never been used before by me or my school and thus presented several unknowns. It was important before presenting to colleagues and relying on it for the main study to check that it produces the necessary results in a functional format and that any unforeseen technical difficulties could be resolved.

Plan 1

In line with the Convention on the Rights of the Child, all pupils taking part in this project were verbally informed of the study, how they would be involved and how the data would be used (UNICEF 2010). Pupils had the option for their data to be excluded but they would still participate in the lessons and assessment task as it is deemed modus operandi and unharmful; none of the pupils in this project opted out. It is possible that pupils felt unable to opt out of the project because of my senior status to them in the school. It is therefore useful when conducting studies with young people to obtain consent from an adult responsible for them. Parental consent was not sought in this project because it was deemed modus operandi and because consent was attained from other responsible figures - the Headmaster and the Central University Research Ethics Committee (CUREC) (see Appendix 3). Seeking consent from persons responsible for young people is also advisable because they may not understand the full ramifications of participating in research, especially if they are vulnerable (BERA 2018). Furthermore, since headmasters have a legal obligation to protect the pupils in their school, informing them allows them to uphold this duty

Request for this consent was emailed to the headmaster as this is the school's typical method of communication and is checked regularly by staff. The email followed the advice from Babione & Anderkin (2015) regarding what content should be included to ensure that they were fully informed about the study (see Appendix 4). The email also included the contact details of my university supervisor in case they wished to raise any concerns or complaints during the research.

Act 1

To avoid advantaging some Year 9 pupils in the main study, Year 7 pupils created the mind maps for the pilot study. Year 7 were chosen because they are a non-exam class of similar size (18 pupils) to my Year 9 class, which would give an indication of the time required to mark the scripts in the main study. Their mind maps were based on shape and space, a topic we were about to start. The assessment took place in lesson time in my classroom with table and chairs set up for exams. Instructions were given verbally. Pupils wrote their names on the back of the mind maps so that their names would not be uploaded to the website and the scripts would be anonymised to avoid bias when marking. The pupils were given as much time as they needed to write down everything they knew about shape and space, this took no longer than 10 minutes.

Observe 1

The scripts were scanned and uploaded to the website. IT support was required to attain the correct format. Once uploaded, marking began very quickly. Figure 7 displays a screenshot of how the scripts were presented on screen. It shows that the previous comparison took 9.4 seconds to judge and the current script is the 25th comparison

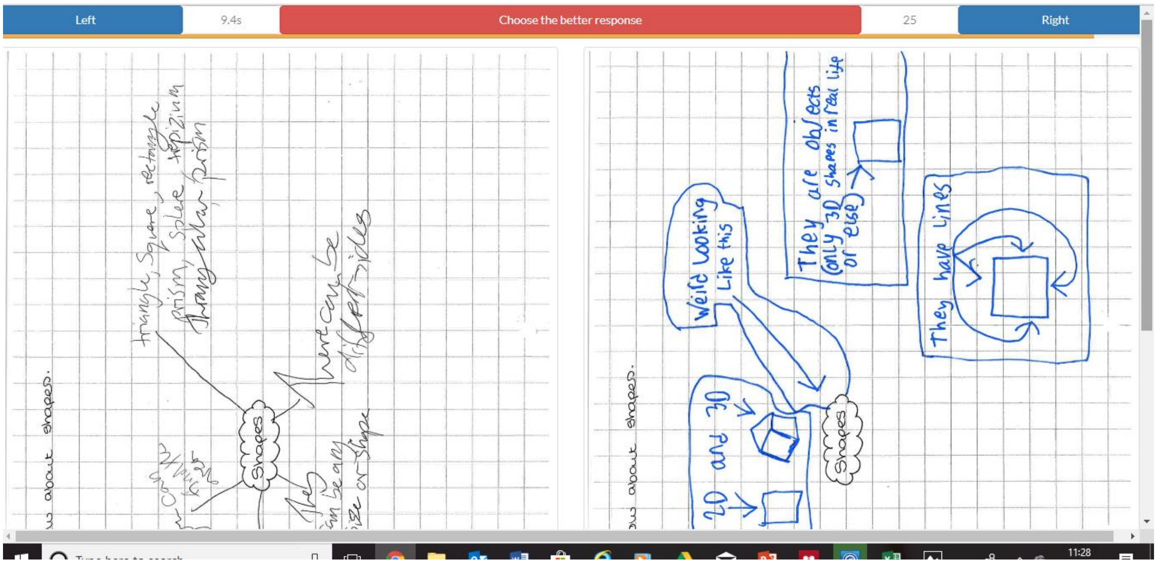


Figure 7. A screenshot of the www.nomoremarking.com from the trial with Year 7

Reflect 1

The trial of the website was very successful. Once the scripts were uploaded, marking began quickly and took very little time as it was often easy to see from a quick glance which script was better. Furthermore, the website generates a lot of statistics downloadable in Microsoft Excel format. All the information needed for this project was available.

A few changes will be made for the main study. All the mind maps in the main study will be on plain paper in portrait form. In the pilot, all the mind maps had a landscape format. This was problematics because the website only presents work in portrait form, which made marking quite awkward. The plain paper will give pupils greater flexibility in presenting their work. Finally, a clear sampling frame will be created for the main study and the codes will be used with the website so that the results can be easily understood and so that pupil names are not uploaded to the website.

Cycle 2

Plan 2

Research Design

A non-randomised quasi-experimental design was used to observe whether the model has an impact beyond typical teaching without the model.

Paired samples t-test

A Shapiro Wilk test was used to check that the data follows a normal distribution. As the data was normally distributed a t-test was conducted according to the following hypotheses:

H_0: There is no difference between the before and after average scores $(\mu_a = \mu_b)$.

H_1: There is a difference between the before and after average scores $(\mu_a \neq \mu_b)$.

Description of experimental and control groups

The groups were non-randomised, consisting of existing sets based on attainment in school tests. Using non-randomised groups can make them less comparable and diminish the validity of results. However, this compromise minimises disruption to me, my pupils and colleagues. Compromises like this are important to take in practitioner research in order to maintain professional relationships (Wilson 2017; BERA 2018). Not using mixed ability classes for the study also best replicates typical practice at my school.

My Year 9 Set 2 class were selected as the experimental group for the following reasons. Year 9 were initially selected as the most appropriate year group because they would not be taking national exams this year and are the biggest which presents a larger pool of pupils that could be involved in the study. Set 2, the class I teach, are also typically well engaged and responsive, which are significant factors in concept development (Coutinho 2014; Skemp 1976; Klausmeier 1992). They are a middle set (4 sets in total) giving the options of Set 1 and Set 3 as preferable control groups.

The control group were selected from the sets of Year 9 mathematics teachers who volunteered to assist. The email asking for volunteers can be seen in Appendix 5 and followed the guidelines as set out by Babione & Anderkin (2015). Volunteers were sought because Wilson (2017) and BERA (2018) believe it is important to treat colleagues with dignity and respect. Fortunately, the volunteers included the teacher of Year 9 Set 1. If Set 2 pupils manage to make more progress than Set 1 pupils, it is a powerful indicator of the efficacy of the model.

Sampling Frame

A sampling frame was created so that pupil names would not be entered onto the www.nomoremarking.com website. This eliminated the risk of their data leaking. Furthermore, it meant that teachers could judge the work without bias towards a pupil. Whilst it is true that some handwriting or content could have made the pupil identifiable to their teacher, a sampling frame minimises this risk. Each script was coded such that a script labelled 01A and 01B would be the pre- and post-intervention scripts of pupil 1, respectively. Initially the Excel spreadsheet linking the code to the pupil's name was kept on a memory stick, as approved by CUREC. However, with the introduction of the General Data Protection Regulation in May 2018, the school forbade keeping data on pen drives and therefore it was moved to the school's secure system in March 2018 (EU 2016). This data will be deleted in December 2018.

Choice of topic for the intervention

The Frayer Model can be used with any subject and any topic (Dunston & Tyminski 2013). Averages was the topic chosen as it was next on the scheme of work and, again, would cause the least disruption to normal ways of working.

Vocabulary Selection (Collaborative Activity)

Colleagues were consulted on the vocabulary that should be used with the model. This approach was also used by Monroe & Pendergrass (1997). By collaboratively selecting the vocabulary, the list of words is extensive but moderated and appropriate. The activity was simple; in pairs, colleagues created a mind map featuring vocabulary considered relevant to Year 9 pupils studying averages. An example of a response can be seen in Figure 8.

Ethical considerations were made when organising the activity. It took place in one of the weekly departmental meetings. A separate meeting would have been difficult to arrange as many in the department have dual roles. This reduced an additional burden on colleagues and by making it easy for them, it would encourage participation. My colleagues were informed of the proposed activity beforehand and offered the choice to opt out (see Appendix 6). Unfortunately, it would be obvious who opted out as they would physically have to leave the meeting for this activity. This was not ideal as BERA (2018) suggests that some colleagues may feel obliged to participate and therefore uncomfortable about opting out in front of others. Therefore, the activity was completed in pairs to alleviate individual performance pressure. This issue was especially pertinent to this activity as the Head of Mathematics and Deputy Head Academic were present. The teacher of the control group was excluded from this activity over concerns that greater awareness of the topic's key vocabulary may affect their teaching.

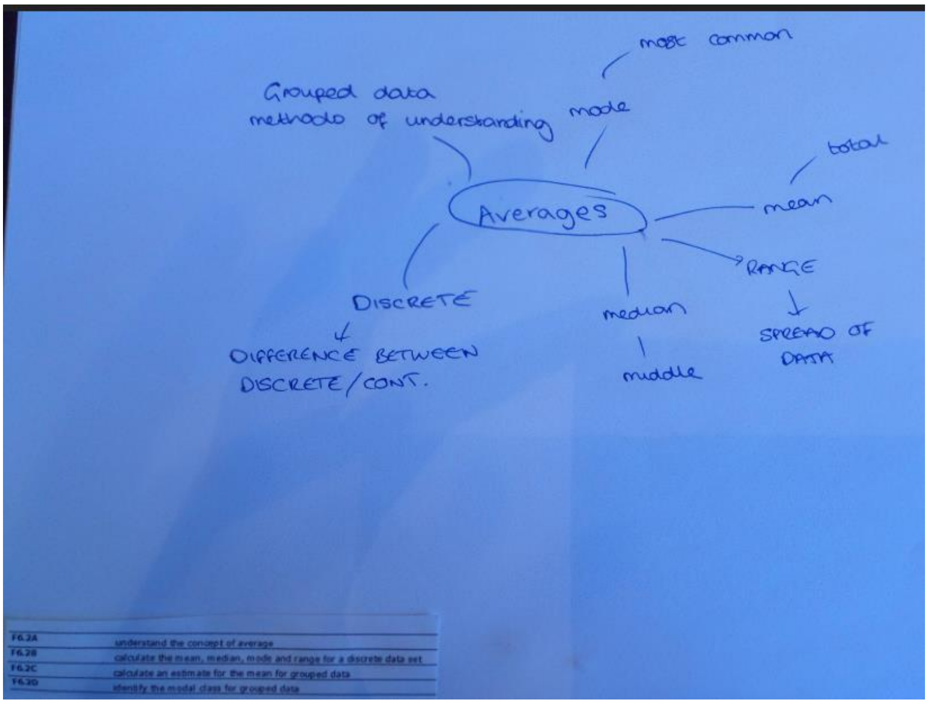


Figure 8. Showing the vocabulary deemed by the department as important for Year 9 pupils to know when studying Averages.

The intention was to use all the vocabulary suggested from this activity. Upon reflection it seemed that some of the more everyday words would be too simple and that using the model for all of them may lead to disengagement. Therefore, the model was only used with the technical vocabulary. The activity was still useful as I personally would not have considered some of the words offered.

Lessons

When not using the Frayer Model, the lessons included the teaching resources I typically use for this topic. The model was implemented in different ways during the intervention to reflect the prior knowledge of the pupils and my style of teaching, as did the teachers in Adams' (2010) and Monroe & Pendergrass' (1997) study. But effort was made to ensure that pupils were engaged in generating the content of the model in some way. This seemed an appropriate choice since I am aware of the benefits when pupils generate their own content (Zandieh & Rasmussen 2010; Watson & Mason 2002). Photocopies of the model were therefore made in advance of the lessons so that they were available for pupils to use and stick into their books.

Assessment Task

All Year 9 classes were held at the same time making it challenging for me to administer the assessment task to the control class. As the task was simple, the teacher of the control group led it instead. They were instructed that it must be done in exam conditions and that there was no time restriction so that pupils would have time to communicate everything they could recall about the concept. Although many pupils may have seen mind maps before, instructions were supplied to all to give clarity and help maintain consistency across the groups (see Figure 9). The instructions encourage pupils to write down relational knowledge along the connecting lines. Pupils were given the opportunity to ask questions before the assessment but were not offered help during the task. I cannot confirm for certain that this was the case for the control group.

Your task is to write down everything you know about averages using a mind map. You should try to include examples, diagrams, definitions etc. to show your understanding of the topic. Along the connecting lines in the mind map, you should explain how the two nodules are related.

You have as much time as you need to complete the task.

You must only write your name on the back of the mind map.

Figure 9. Shows the instructions provided to the pupils in both groups.

Act 2

Lessons

The intervention lasted for one week, equating to four one-hour lessons. A Frayer Model was created for each of the following concepts: Averages, Mean, Median, Mode, Continuous Data and Discrete Data. The different ways that the model was used tended to depend on the assumed prior knowledge of pupils and thus whether they would be able to create content. For instance, averages are a commonly used term outside of mathematics lessons that pupils are likely to have previously encountered. Therefore, prior to teaching the topic, the model was projected on the board and pupils contributed to the different quadrants in a whole class discussion. Some pupils had misconceptions about averages, suggesting rounding, the nth term and standard form as examples. After discussion these were moved to the non-example quadrant. The model was completed again by pupils in pairs for mean, median and mode as a reflection activity after teaching further content. The model was also used to introduce the categories of discrete and continuous data; I gave some examples and described the terms, after which pupils made further contributions. Photographs of the completed Frayer Models for each of the examples described above can be seen in Appendix 7.

Assessment tasks

The pre-teaching assessment took place in the lesson before the start of the intervention and the post-teaching assessment in the lesson after teaching had finished. All pupils finished their mind maps in less than 10 minutes.

Observe 2

The results of any pupils that missed either of the assessments or any lessons during the intervention were excluded. This is because the intervention was so short that missing a single lesson might have made a significant impact. Table 2 shows the final number of pupils that took part in Cycle 2, alongside some characteristics that indicate the resemblance of the groups.

With each pupil completing a pre- and post- teaching assessment, a total of 70 scripts were marked. Six judges, including myself, participated in the marking process completing a total of 720 comparisons to achieve the quantitative results. This took approximately 15 minutes at the end of a departmental meeting. The results have a good level of reliability as only 7 scripts had a high infit score and the website generated a Rasch Reliability rating of .91.

	Experimental Group	Control Group
Total Number of Pupils	15	20
Number of EAL Pupils	1	2
Number of SEN Pupils	9	10
Female/Male Numbers	4/11	5/15

Table 2. Shows some characteristics of the intervention and control group

In addition to the evaluation criteria stated earlier, two further lines of inquiry specific to the topic of averages that arose during the marking process will be explored. Firstly, the model's ability to address the misconception that range is a type of average and secondly a discussion about the absence of discrete and continuous data from pupils' mind maps.

Average progress

Table 3 shows the change in true score and the change in the 0-100 scaled score made by pupils in the experimental and control group. The results indicate that the pupils in the experimental group progressed on average by 17/100 marks whilst those in the control group did not progress at all. There are no outliers in the data which suggests that the unusual outcome of the control group is not a consequence of the extreme results of a few pupils.

Average Progress	Experimental Group	Control Group
True Score	2	0
Scaled Score	17	0

Table 3. shows the average changes in progress in terms of their true score and scaled score

Paired samples t-test and effect size

A paired samples t-test indicates that the progress made by the experimental group was reliable $(\rho = .016)$. The computed effect size suggests the intervention had a large impact on the development of pupils' conceptual understanding $(r = .71)$.

Evidence of the Frayer Model features

The pre- and post-teaching mind maps were searched for direct evidence of the impact of the Frayer Model by counting the number of scripts that offered a definition, facts/characteristics, examples or non-examples. The results are in Table 5. The values indicate the total number of scripts containing these features, not the total number of times that a feature was evident. Distinguishing whether some features were definitions or facts/characteristics was challenging. For consistency, descriptions of how to find the mean, median and mode contributed to facts/characteristics whilst suggesting that the average represents a set of data was classified as a definition. The non-examples were only tallied when pupils explicitly identified them as such. Therefore, it did not include those scripts which included range without further elaboration, of which there were many. Note that prior to using the Frayer Model, the term 'non-example' has not been used with Year 9 Set 2.

Pre - Teaching	Experimental Group (15 pupils)	Control Group (20 pupils)
Definition	0	0
Facts/Characteristics	8	10
Examples	9	5
Non - examples	0	0

Post - teaching	Experimental Group (15 pupils)	Control Group (20 pupils)
Definition	2	0
Facts/Characteristics	13	15
Examples	12	5
Non – examples	3	0

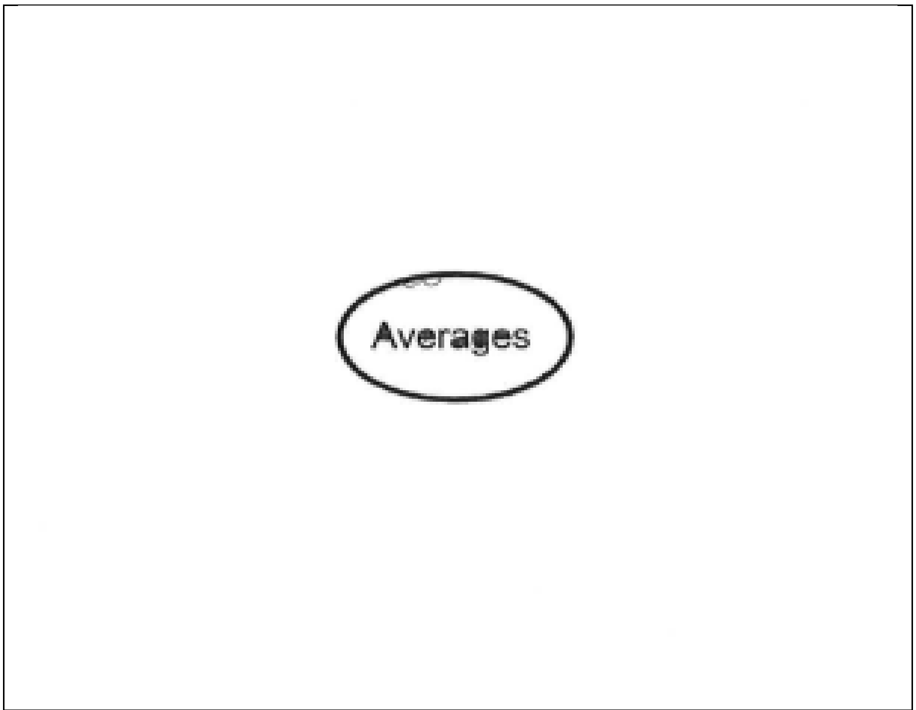
Table 5. Shows how many pupils out of those that took part in the intervention included the categories of the Frayer Model in their pre- and post-teaching mind maps.

Unfortunately, the set sizes are not equivalent, so direct comparisons between the groups can be misleading and should be made carefully. Both groups saw a rise in the number of pupils offering facts/characteristics. However, this was the only difference for the control group, whereas the pupils in the experimental group increased the number of references they made in every quadrant.

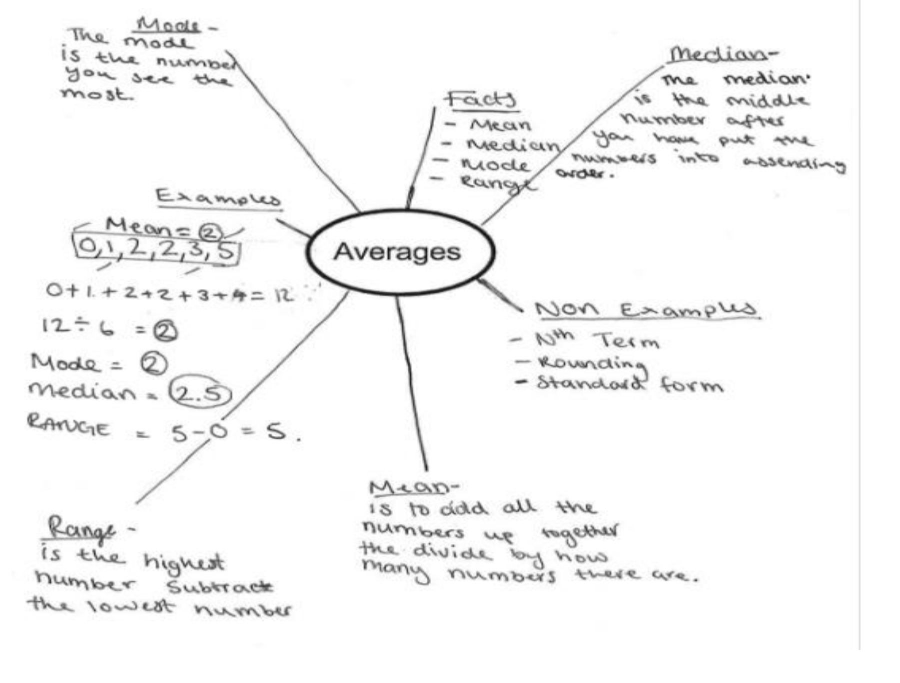
Pupil A - Most progress

The quantitative data indicates that a pupil (Pupil A) from the intervention class made the most progress overall. Their pre- and post-intervention test can be seen in Figure 13. Assuming the pupil engaged with the task, this pupil seemingly had no prior knowledge of the topic and increased their scaled score by 72/100 points by the end of the intervention. The entire structure of their post-teaching mind map seems to resemble the CD-Frayer Model used in Monroe & Pendergrass' (1997) study (see Figure 2) as it also includes sections labelled facts, examples and non-examples. This could be a direct consequence of having used the Frayer Model.

With regards to content, Pupil A is now able to explain how to find the separate averages from raw data and has demonstrated the procedures using their own examples; although there is an error in their finding of the median. Given they correctly described the procedure this suggests it was either a mistake or, as Vinner (2002) suggests, they may have memorised the explanation but not fully understood how to apply the procedure. Pupil A also managed to list the range as a non-example of averages, this suggests that their negative knowledge of the concept has also improved.



Pre – Intervention Test



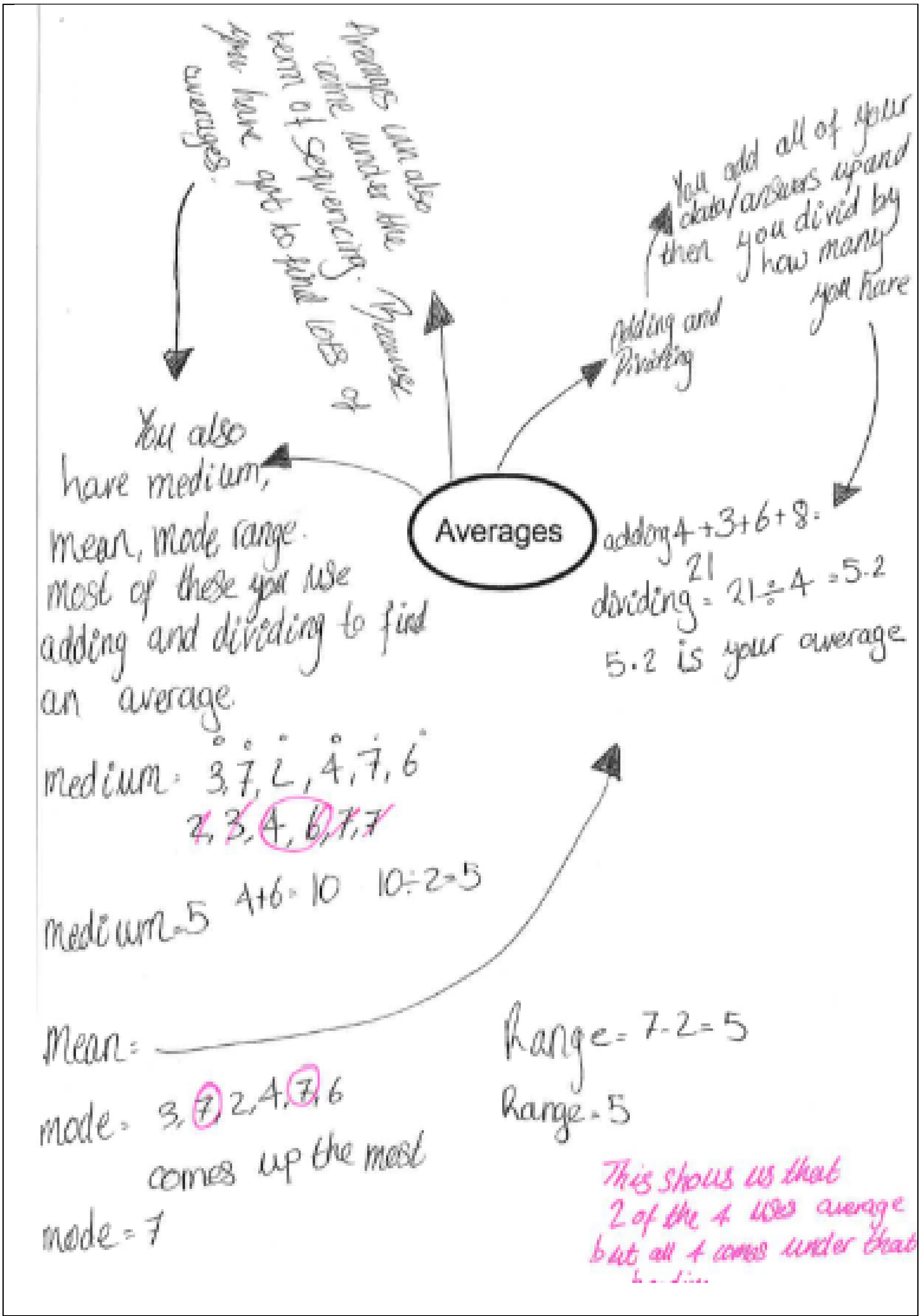
Post – Intervention Test

Figure 13. Shows the pre – and post – test of Pupil A who made most progress.

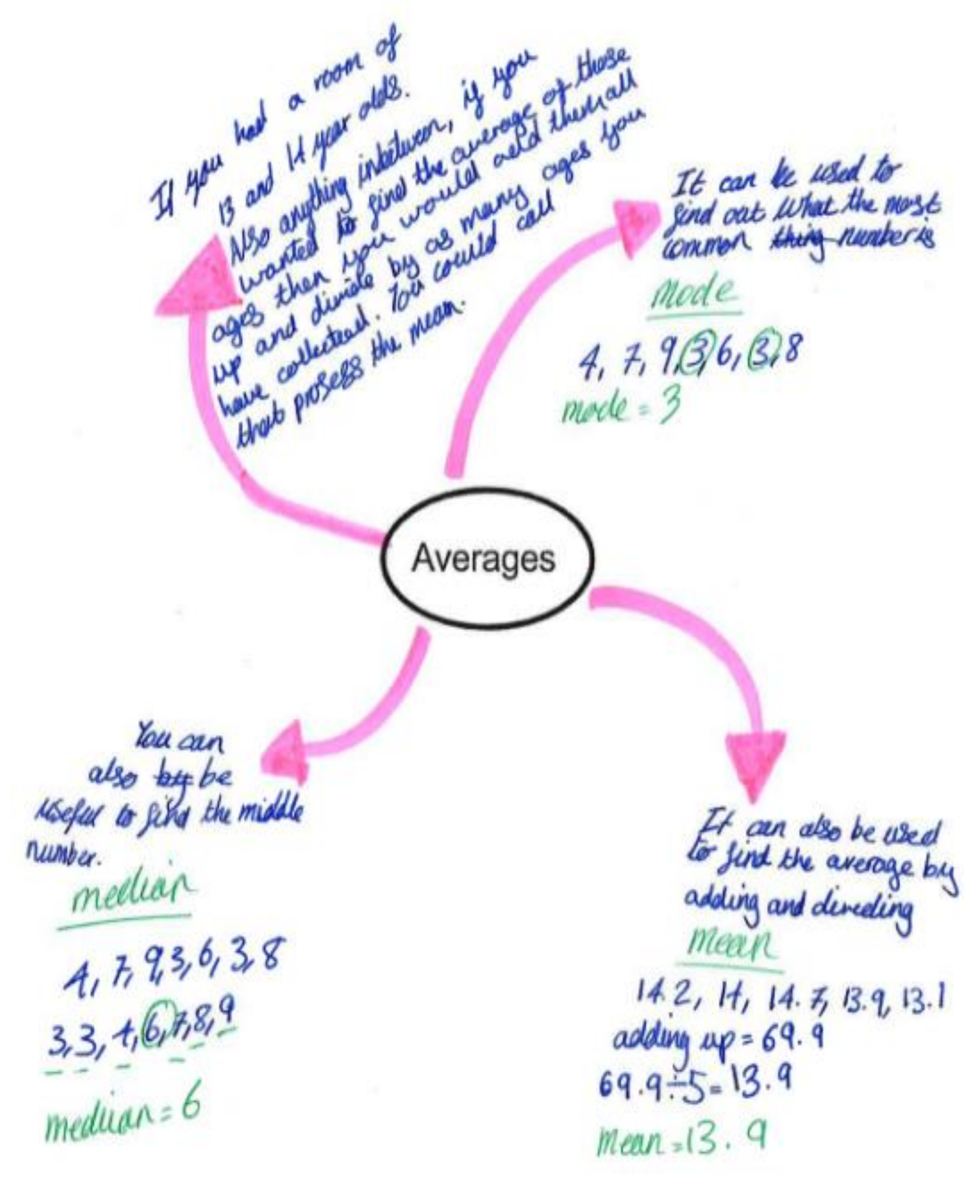
Least progress

The pupil (Pupil B) that overall made least progress was from the experimental group and their score decreased by 70/100 marks. Their pre- and post- tests can be seen in Figure 14. There are two main changes to the content of the mind maps. The pupil's post-script no longer includes the range and has different examples. By not including the range, the pupil has less content on their page which could lead a judge to believe it is not as good as the first script. However, the pupil was in the experimental class where range was highlighted as a non-example and thus it could be suggested that the pupil's understanding has improved because range is no longer written. Perhaps by emitting the range, they are trying to show that they now understand that it is a non-example. Ideally, the pupil would have still written range but explained its relationship to averages. The second major change is that the procedural examples in the post-script now include decimal figures as opposed to natural numbers. This could indicate that the pupil is experimenting with the limits of the concept and is aware that averages can be calculated with decimals. Another possibility is that the pupil simply had a calculator available for the post-test and was therefore able to do more challenging calculations.

The post-script also has a slightly different and seemingly more organised structure than the first. We can only speculate as to the cause of this. It could be that the Frayer Model has helped the pupil to organise their knowledge. But, having recently revisited the subject, the content may be easier to recall and organise neatly. Another possibility is that the pupil can organise it better because this is their second attempt at creating a mind map.



Pre – Intervention



Post – Intervention

Figure 14. Shows the pre- and post-intervention scrip of a pupil whose score decreased over the intervention period.

Addressing misconceptions

It was noted when glancing through the pre-teaching mind maps that many pupils identified the range as a type of average. This was also noticed and addressed when using the model in lessons. Since the literature implies that the quadrants in the model may help to clarify concepts and prevent misconceptions, it is of interest to see how effective the model was in achieving this. Table 6 shows the number of mind maps referring to range as an average before and after the intervention. There seems to be little difference in either group in the number of pupils with this misconception. However, there were three post-intervention mind maps that identified the range as a non-example of the average. These were not included in the figures below.

	Experimental Group (15 pupils)	Control Group (20 pupils)
Before	7	17
After	6	17

Table 6. Displays the number of scripts containing the term 'range'

Discrete and continuous data

Sometimes reflecting upon what is absent can bring another perspective and additional insight. Colleagues identified discrete and continuous data as being terms that Year 9 pupils studying averages should be familiar with. Both terms were therefore introduced using the Frayer Model. However, none of the mind maps contained these terms.

Reflect 2

The aims will now be reviewed with respect to the observations collected in Cycle 2.

1) To find a way of observing the development of conceptual understanding

Mind maps were used to try and gain insight into a pupil's concept image of averages. From a teacher's perspective, they required few resources and little time and skill to prepare. They were completed quickly by pupils, only taking at most 10 minutes of a lesson. Also, as hoped the pupils seemed to use the flexibility inherent in mind maps to draw frequency tables, create examples and write single words or paragraphs (Abi-El-Mona & Adb-El-Khalick 2008; Price & van Jaarsveld 2017). This provided different perspectives, revealing more about the pupil's level of understanding; such as when Pupil A stated a correct method for finding the median but could not apply it correctly in their example.

There were concerns regarding the mind maps of some pupils that were sparse. It is possible that the pre-teaching mind map of Pupil A was empty for reasons other than their lack of prior knowledge. For instance, they may not have engaged with or understood the task, or they may have struggled to communicate their understanding via the mind mapping format. If this is the case, the results may have been skewed and the progress of this pupil exaggerated.

It could be argued that using an assessment task with which pupils are already familiar with would have prevented any teething issues. However, as this project needs to measure conceptual understanding rather than general attainment, this would not be ideal because having the additional stimuli of the context of the question may limit a pupil's potential to express their knowledge of the concept (McClure et al. 1991). Or it could help a pupil too much. The concept image comprises all the knowledge that one associates with the concept, so the concept's name should be sufficient stimuli (Tall & Vinner 1981). To assist pupils with understanding the assessment task, the instructions could be reviewed. This will be done for Cycle 3 because it may also solve another issue that will now be raised.

Not all pupils used the mind maps as intended and this may have limited their potential to show a pupil's concept image. Freeman et al. (2003) and McClure et al. (1991) believe that the structure of mind maps can be analysed to offer additional insight. This was true to an extent in Cycle 2. For instance, a clearly structured mind map may reflect a pupil's clarity about the structure of the concept and the ideas they associate with it. However, this is pure speculation. It seems difficult to interpret the structure of mind maps beyond speculating what it could mean with regards to a pupil's concept image. This issue may have been a consequence of pupils not labelling the connections between nodules in their mind maps. Labels would have aided clarity. Therefore, the instructions regarding the construction of the mind maps for Cycle 3 were changed in the hope that it would encourage more pupils to label their mind maps fully.

The changes in the structure of Pupil B's mind maps provides evidence that using a marking rubric, like those mentioned in Evrekli et al.'s (2010) paper, could have been misleading. There are very few changes in the content between the before and after mind maps offered by Pupil B. However, the structure of this content has changed. The pupil seems to have a more consistent way of organising the content in the post-teaching assessment which involves fewer drawn connections. According to the marking rubrics, this would result in a lower score even though the content is comparable. Thus this provides reassurance that not applying the commonly used evaluation of rubrics was suitable for this study.

In Cycle 3 the assessment task will also be conducted slightly differently. Due to timetabling limitations, the assessment tasks for both groups in Cycle 2 were conducted by different people. Instructions were given to the teacher to try and ensure the groups had the same experience. As the task was simple, it was a compromise that solved a tricky logistical situation and was unlikely to have a large impact on the reliability and thus validity of the results. This is an instance of when the rigour of a study is compromised due to the nature of practitioner research. It was not ideal as the teacher may have not followed the instructions given, or their implementation of the instructions may have differed from mine. Timetabling changes made it feasible for me to conduct both the assessments in Cycle 3. Some may argue that as the researcher, I may be biased and that neither of the tests should be conducted by me, however awareness of the potential for this bias should help to minimise this risk.

Using the comparative judgement technique in conjunction with the www.nomoremarking.com website was a successful collaboration task. The website made the process very easy, the judges could follow a link and start marking almost immediately. There were no technical issues that needed to be resolved during the marking session. Each judge finished their allocation of 120 comparisons quickly so that comprehensive quantitative data was achieved in less than 15 minutes of marking. The judges seemed to enjoy the novelty of this new approach to assessment in mathematics. There is also a possibility that their practice may have benefitted by reflecting on pupil's work. Although collaborative marking reducing the workload, it also helps to reduce the bias caused by having only one judge. The reliability and infit scores show that the scripts were marked with high levels of consistency. This may be expected as all the judges are mathematics teachers in the same school with the same culture. There is therefore still the potential for systematic bias (Jones & Inglis 2015).

This may mean that the results specifically reflect the department's learning goals for our pupils.

There were two collaborative elements with colleagues in Cycle 2; the process of vocabulary selection and the marking of the scripts. The effectiveness of a collaborative task can be impacted by how members work together. Leading these tasks was atypical of my role. Within the department, I have the lowest seniority as all the other members have dual roles. The unusual dynamics may have impacted the success of these tasks, particularly in the vocabulary selection where there was a lot of interaction within the group. For instance, the Head of Mathematics may have limited their input to allow me to take the leader role.

The impact of roles, I believe, was less significant in the marking of the scripts as it involved very little interaction. After emailing out instructions with the website link and booking the computer room, members of the department marked their allocation independently. Due to the structure of the meeting, arranged by the Head of Mathematics, colleagues could leave the meeting after they had finished their allocation. If colleagues were keen to leave, this may have resulted in the lack of discussion. This was a shame in many ways as discussions about vocabulary and what teachers valued on the mind maps could have been an opportunity for professional development.

A consequence of using a numerical scoring system to measure conceptual understanding is that it can make the results difficult to interpret. For instance, on a spelling test it seems clear that if a pupil scores highly that they have demonstrated some skill with spelling. It is also clear what a model, correct spelling test would look like. However, because of the more complex nature of conceptual understanding and because a marking criterion was not used, it is not possible to have a model, correct mind map. That is, it is not possible to necessarily visualise the mind map of a high scoring pupil or what progress might look like. This is not necessarily a drawback of the method, it is just a consequence of this holistic approach to assessing conceptual understanding.

The thematic analysis of the mind maps worked well in terms of providing more clarity and meaning of the scores. It reveals more detail about the pupil's concept image, especially when the before and after scripts can be viewed together. The analysis showed how pupils refined their definition or explored more of the concept's boundaries. It has also been useful to compare with the scoring system and indicates that using scores alone could have been misleading. For instance, with the pupil whose score indicated their understanding of averages had regressed. As a practitioner-researcher who has additional knowledge about the pupils and the lessons taught, this score seems unfitting. Having this insight because of triangulating the evaluation methods raises concerns about making conclusions based solely on the mind map scores.

2) To observe how the Frayer Model impacts a pupil's understanding of a concept

The quantitative data implies that the Frayer Model had a positive impact on the concept image. The calculated effect size suggests that those in the experimental group made substantially more progress than the control group and this was achieved with a high level of significance. That is, it is unlikely that the greater progress occurred by chance and is likely a result of the intervention. The quantitative data also suggests that the pupils that made the most and the least (negative) progress were part of the experimental group, but the score allocated to the latter is in question.

There is evidence to suggest that the improvement in progress was a result of the Frayer Model. Definitions, facts/characteristics, examples and non-examples are features of the model and occurred more frequently in the scripts of pupils from the experimental group. As well as being an indicator of the impact of the model, these features are indicative of emerging conceptual understanding. Klausmeier's (1992) highest level of conceptual understanding requires pupils to identify examples and definitions of a concept (see Appendix 1). That is, more pupils in the experimental group were offering features that indicate high levels of conceptual understanding.

The Frayer Model was not a complete success as none of the mind maps referenced continuous and discrete data. Both terms were introduced using the Frayer Model (see Appendix 6). Unless the vocabulary selection made by me and my colleagues was misguided, comprehending these terms should not have been beyond their current capabilities. This perhaps indicates that the model is not an effective way of introducing concepts and is best used as a reflective activity when pupils already have some prior knowledge. Another option is that the connection between continuous and discrete data with averages was not made clear. It seems upon inspection of the photograph in

Appendix 6 that this may have been the case. If so, this emphasises the importance of making the relationships between concepts explicit to pupils.

A flaw in the research design means that it is not possible to attribute the success of the experimental group to the Frayer Model. The purpose of the quasi-experimental design was to show the impact that the model makes above normal teaching of a topic. However, as a different teacher taught the control group, this cycle can only indicate that my teaching with the Frayer Model appears to be more effective at developing conceptual understanding than the control teacher's normal teaching. It is impossible to distinguish whether the results reflect the use of the model or, for example, our different teaching styles and resources. Hence, the data from Cycle 2 will not indicate whether integrating the model into my everyday teaching will be an improvement. Adjustments to timetabling were made in Cycle 3 to allow this comparison.

It also worth noting a secondary factor that became apparent during the intervention. This factor was not specifically mentioned as impacting the development of conceptual understanding in the literature, but BERA (2018) and Reeves et al. (2010) suggest that researchers should be aware of how relationship dynamics may influence interventions. This intervention required pupils to create their own content and participate in class and partner discussions which is not a typical feature of our lessons. It is possible that the class dynamics needed to change to facilitate these activities. As such, the efficacy of the model may be limited by the established culture of the class, or its ability to adapt. Additionally, it then seems that the model could become more effective as pupils become used to working in this way, and that the intervention may have a lasting change to the relationship dynamics of Year 9 Set 2.

Cycle 3

Plan 3

Research Design

The structure of action research allows changes to each cycle so that teaching practice can be continually improved (Wilson 2017). This cycle will aim to improve the research design to obtain more useful results that can then be used to inform my teaching practice. To try and isolate the impact of the Frayer Model, both the control and experimental group will be taught and assessed by me. Whilst other factors, such as the participating year group and topic to be taught, will try to be consistent with Cycle 2.

Wilcoxon Signed Rank Test

The Shapiro Wilk test suggested that the data was not normally distributed and thus a Wilcoxon Signed-Rank test was conducted with in this to determine the significance of the results. The test was conducted according to the following hypotheses:

H_0: There is no difference between the before and after scores ($\mu_a = \mu_b$).

H_1: There is a difference between the before and after scores ($\mu_a \neq \mu_b$).

Consent

The new research design creates more disruptions for the school and pupils and thus additional consent was required. To teach both the control and experimental group, I would need someone to cover my Year 9 Set 2 class whilst I taught the experimental group. Doing a 'teacher swap' within the department would enable my class to be taught by a mathematics specialist, reducing the risk of them being disadvantaged by the process. Consent for the swap was sought via email from the Head of Mathematics and the Deputy Head Academic (see Appendix 8).

Teacher Swap

It would have been possible to include two different classes in Cycle 3. This would have prevented Year 9 Set 2 from participating in both cycles, which BERA (2018) warns against doing. However, this would have created further disruptions to timetabling and left Set 2 without their usual teacher for longer. As a compromise, Set 2 were the control group in Cycle 3 where the only change to normal teaching is the use of the assessment task. Set 3 were selected as the experimental group because their teacher (Teacher A) volunteered to participate and because their attainment is more comparable with the control group than Set 4. The letter asking for volunteers can be seen in Appendix 9 and again followed the guidelines set by Babione & Anderkin (2015). Because of their act of good will, this teacher was given a small gift upon completion of the study; it was not advertised as a reward for participating. Table 11 shows the sets and their teacher for each cycle in this study.

	Control Group	Experimental Group
Cycle 2	Year 9 Set 1 (normal class teacher)	Year 9 Set 2 (me)
Cycle 3	Year 9 Set 2 (me)	Year 9 Set 3 (me)

Table 11. Shows which sets were involved in the project and who taught them for each cycle.

An issue with making Set 2 the control group is that they will not be taught using the Frayer Model. They seemed to benefit from the intervention and therefore it could be viewed as wrong to not use the model with them when teaching the next topic. Some may believe that it is removing a learning opportunity. However, the research design flaws mean that the results of Cycle 2 were not conclusive. Therefore, although it is perhaps not ideal to now use Set 2 as the control group, it seems more ethical to properly trial the model before integrating it long term into my practice.

The study was designed so that teaching of the experimental group happened after the control group. This is because the aim of this cycle was for the control group to experience my typical teaching of the topic. Teaching with the Frayer Model first could have influenced my 'normal' teaching. This may have already happened having used it in Cycle 2, but this approach will help to minimise its influence.

Choice of topic for the intervention

To produce results comparable with Cycle 2, it would require teaching Averages again to different pupils of a similar nature. This is not possible in my environment. Venn Diagrams and Set Notation was selected instead for teaching in Cycle 3 because it is a similar, statistics-based topic that Year 9 pupils still needed to be taught. In my absence from teaching Set 2, Teacher A taught them another topic from the syllabus. Verbal permission to alter the order of topics to be taught was gained by the Head of Mathematics. This was done verbally because this is our typical form of communication.

Vocabulary Selection (Collaborative Activity)

To reduce the number of non-technical words offered by my colleagues, more verbal explanation of how the vocabulary would be used was given at the beginning of this activity. An example of one resultant mind map can be seen in Appendix 10. All members of the department, including the teaching of the control group from Cycle 2, participated in the activity. Arguably I should not have been present as it may have influenced my teaching of the control group.

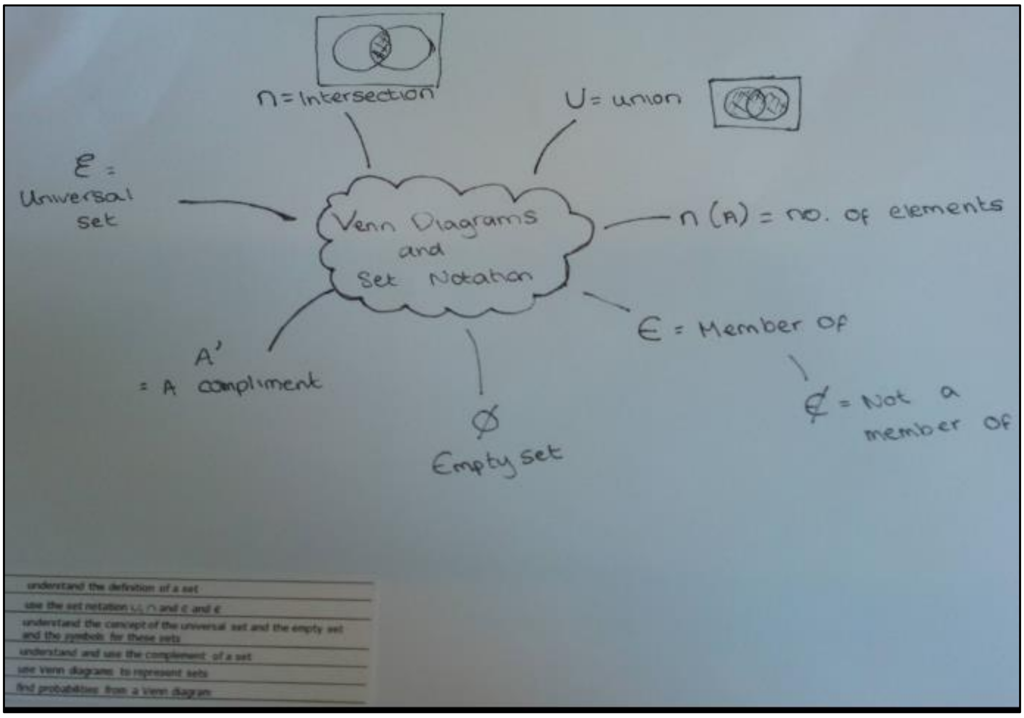


Figure 15. Mind map of the vocabulary identified by the Mathematics Department as significant for Year 9 pupils to learn for the Venn Diagrams and Set Notation topic.

Assessment Task Instructions

As pupils did not label their mind maps as desired in Cycle 2, the task instructions for pupils were reviewed for Cycle 3 and can be seen in Figure 16. The only change being the addition of a diagram to exemplify the instructions. Due to time constraints, the instructions were not trialled before using in Cycle 3.

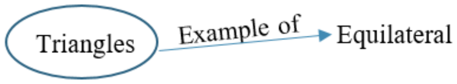
> Your task is to write down everything you know about averages using a mind map. You should try to include examples, diagrams, definitions etc. to show your understanding of the topic. Along the connecting lines in the mind map, you should explain how the two nodules are related.
>
> 
>
>
> You have as much time as you need to complete the task.
>
> You must only write your name on the back of the mind map.

Figure 16. Copy of the task instructions given to all participating pupils in Cycle 3.

Act 3

Lessons

The Frayer Model was used for all the words displayed in Figure 15. Lesson plans were used to try and ensure consistency in teaching between the groups, like in Monroe & Pendergrass' (1997) study. Whilst pupils in the experimental group used the model, the control group spent more time practising past exam questions on the topic. In all instances, the model was used as a reflective activity after teaching of content because it became apparent that pupils had little prior knowledge they could contribute. Pupils engaged well with making their own content for the model but struggled to think of non-examples. I too found it challenging to offer non-trivial non-examples for the terms in this topic. It should be noted that the experimental group are usually supported by a teaching assistant who was present for the whole of the intervention.

Assessment

Due to unforeseen circumstances, the test for the control group could not be administered until a month after finishing the teaching of Venn Diagrams and Set Notation. This was because of changes to whole school events in the week before the Easter holidays. Fortunately, it was possible to also wait a month between the teaching period and assessment for the experimental group. This was different from Cycle 2, where the assessments were conducted in the lesson after teaching finished.

Observe 3

A total of 31 pupils participated in Cycle 3. Some characteristics of the groups can be seen in Table 7. Except in gender ratio, they seem comparable.

	Experimental Group	Control Group
Total Number of Pupils	15	16
Number of EAL pupils	3	2
Number of SEN pupils	10	10
Female/Male	9/6	5/11

Table 7. Indicates the nature of the groups involved in the study according to gender, EAL and SEN.

Six judges from the department helped to mark 62 scripts. Each judge was allocated 105 comparisons which took no longer than 10 minutes to complete. As one colleague unforeseeably could not attend the marking session, fewer comparisons per script were made in this cycle. This explains the slightly lower but still very high Rasch reliability rating of 0.88.

The repeated involvement of Year 9 set 2 allows for comparisons between the scripts of pupils that took part in both cycles. These could show how an individual responds to the assessment task differently because of the Frayer Model. The scripts of two of these pupils will be randomly selected and presented. Random selection will be used to give a wider perspective of the scripts as only the more extreme cases of pupils that have made the most and least progress have been analysed in depth so far.

Average Progress

The average progress according to the generated true and scaled scores can be found in Table 8. It indicates that the experimental group made more progress in their conceptual understanding; on average progressing by 3/100 as opposed to 2/100 in the control group. The true to scaled score ratio suggests a greater range in the quality of scripts in this cycle than in Cycle 2.

Average Progress	Experimental Group	Control Group
True Score	27	16
Scaled Score	3	2

Table 8. Average progress made by the experimental and control group according to true score and scaled score.

Wilcoxon Signed Rank Test and effect size

The Wilcoxon Signed Rank Test indicates that the progress made by the experimental group occurred with a high level of significance $(\rho = .000027)$. The calculated effect size suggests that the intervention had a large impact on pupils' development of conceptual understanding $(r = .96)$.

Evidence of the Frayer Model features

The results of checking the scripts for features of the Frayer Model can be found in Table 10. Deciding what constitutes a definition rather than a fact/characteristic for this topic was challenging. Isolated notation such as '∩' and/or the word 'intersection' would count as a fact/characteristic of set notation, whilst descriptions of their associated meanings would constitute a definition. Typically, the examples were simple Venn diagrams with two intersecting sets that were sometimes based in a context. Initially the figures were similar and although both groups made improvements in three of the categories, the experimental group made more.

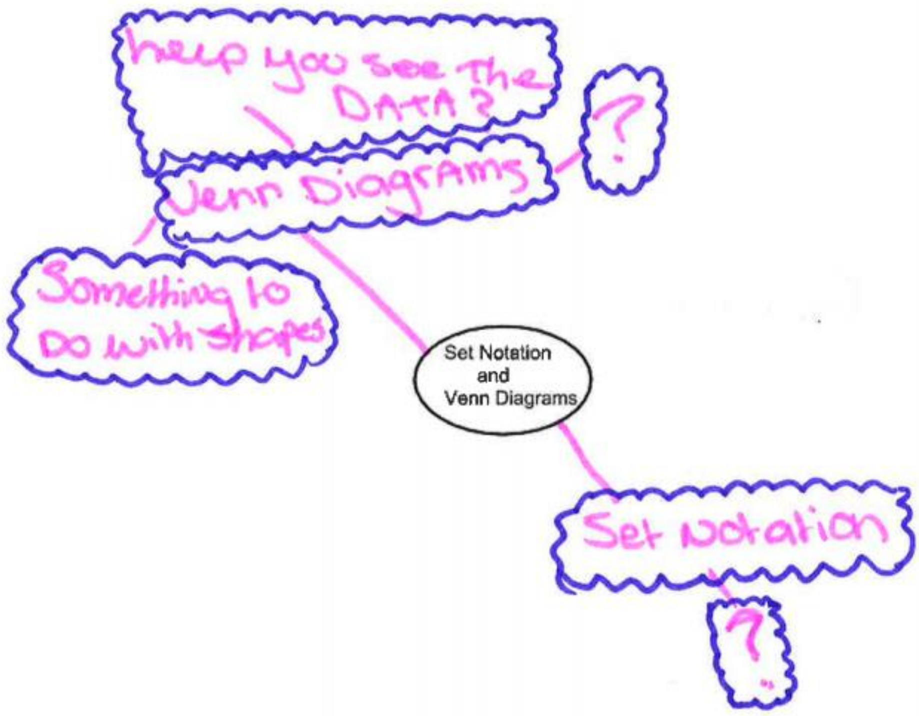
Pre – Teaching	Experimental Group **(15 pupils)**	Control Group **(16 pupils)**
Definition	0	0
Facts/Characteristics	5	5
Examples	10	8
Non – examples	0	0

Post – Teaching	Experimental Group **(15 pupils)**	Control Group **(16 pupils)**
Definition	8	4
Facts/Characteristics	10	8
Examples	14	13
Non - examples	0	0

Table 10. Shows the number of pre- and post-teaching mind maps showing features of the Frayer Model

Most progress

A pupil from the experimental group made the most progress, increasing their score by 66/100. Figure 17 suggests that the pupil initially had little knowledge of the topic. The post script shows that they have learnt some notation and now understand the meaning of different regions in a Venn diagram. The pupil has also exemplified a standard exam-style set notation question, but not proceeded to answer it.

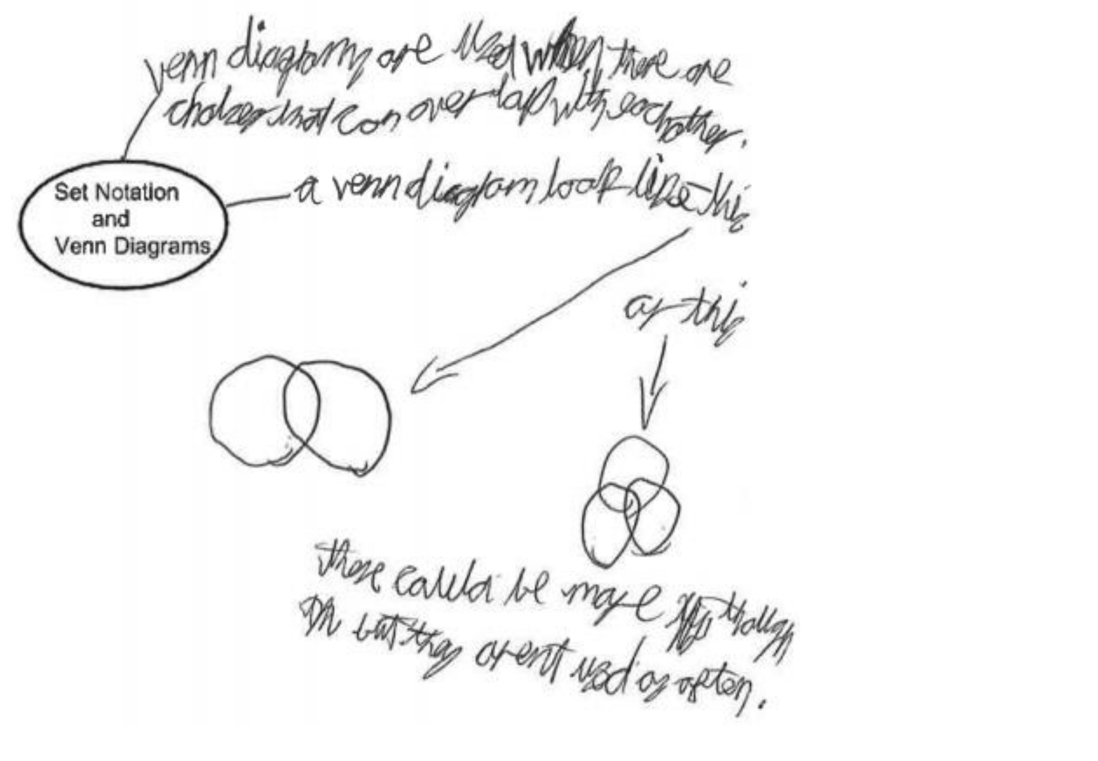

Pre – Intervention

Post – Intervention

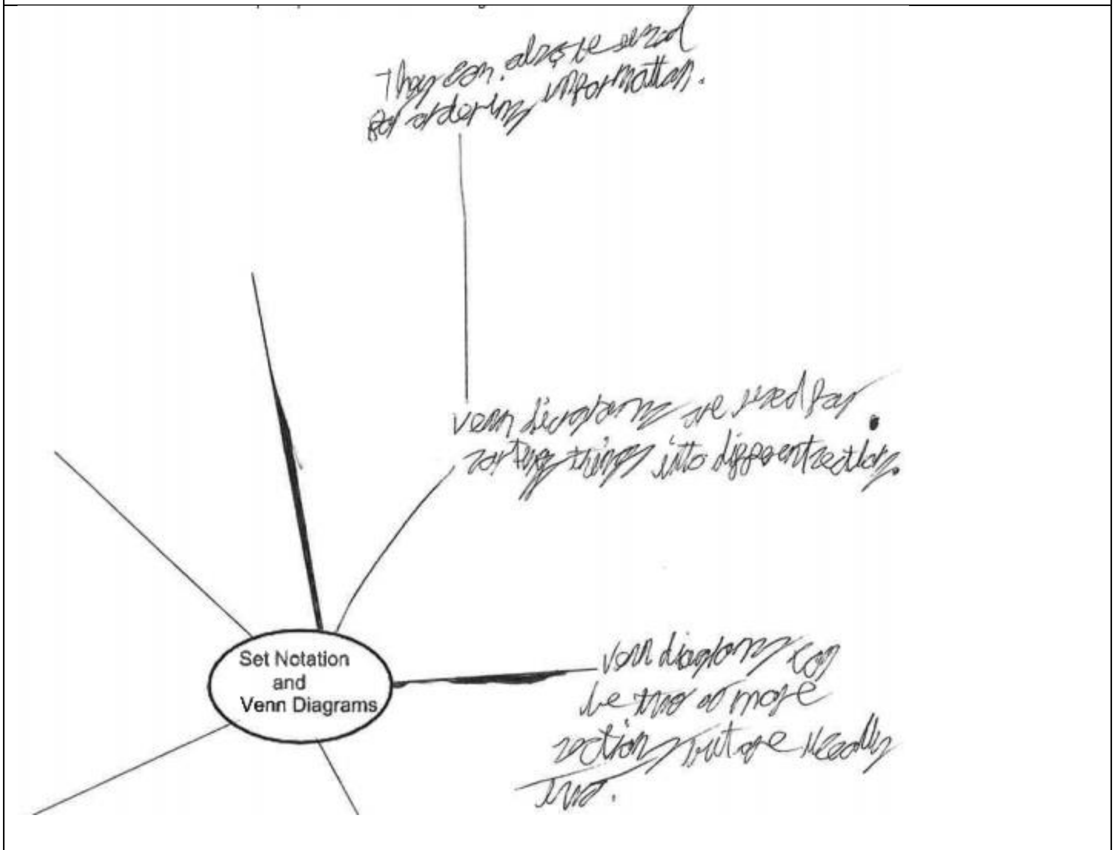
Figure 17. Shows the before and after scripts of the pupil that made the most progress.

Least progress

Two pupils, one from each group, made the least overall progress which was a decrease of 8/100 marks. Figure 18 shows the script of the one that used the Frayer Model. This pupil initially had some prior knowledge of the topic, but there is no apparent change in the content of the post-teaching script except a change in their format of communication; providing written explanations instead of diagrams. It seems that the pupil has not progressed, but not regressed either as the score would suggest.



Pre – Teaching Script



Post – teaching Script

Figure 18. Shows the before and after script of a pupil from the control group that appeared to decline in progress.

Comparison of pupils from Cycle 2 and Cycle 3

A total of 11 pupils from Year 9 Set 2 participated in Cycle 2 and 3. Unfortunately due to changes in the study design, these pupils had already encountered the Frayer Model before participating in the control group which may taint their experience of the typical teaching in Cycle 3. Pupils C and D were randomly selected and all four of their scripts can be seen in Figures 19 and 20, respectively.

The examples Pupil C gave in Cycle 2 were very similar across both assessments, with both including five single digit figures starting with 1, 1. There were some changes over the intervention period. They acknowledge range as a non-example of an average and can now demonstrate the procedure for finding the mean from grouped data. These new features have been added without changing the initial structure of the mind map and have not been connected to the existing content.

Pupil C's scripts in Cycle 3 (without the Frayer Model) seem quite different. The pupil has less prior knowledge about this topic. There is more content in the post-test, but it does not appear as well explained or as cohesive as in Cycle 2. Furthermore, as 'Set Notation' has been separated from the examples of notation, it could suggest that the pupil has some confusion about what it is they have learnt. Due to the use of different topics between cycles in cannot be concluded that the lack of clarity is a result of having not used the Frayer Model for this topic.

In contrast, Pupil D had little prior knowledge of both topics. This makes it easier and more viable to compare how their concept image developed across the cycles. In both cycles, the pupil offered new technical content and examples in their post-teaching scripts. However, the example they gave in the post-teaching Cycle 3 mind map of the union of sets is wrong and does not align with the correct description they gave. On the other hand, the post-intervention script does not contain any errors. It also has a clearer structure with appropriately labelled connections and since they recognised that the range is a non-example of an average, it insinuates the pupil has developed negative knowledge of the concept too.

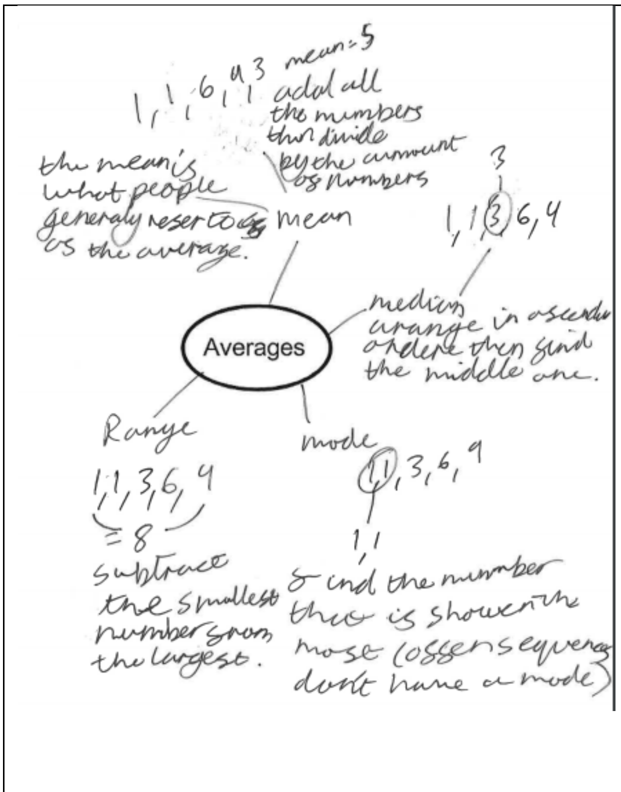
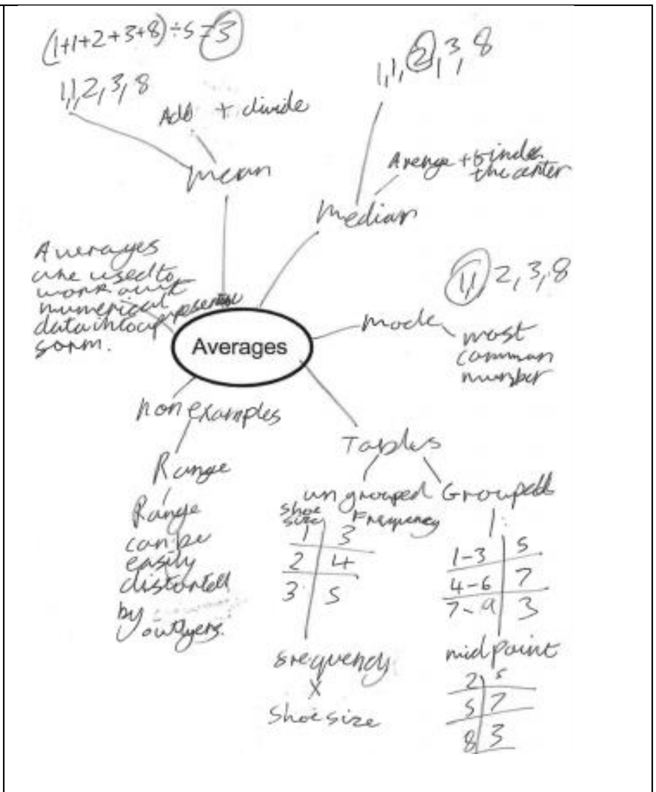
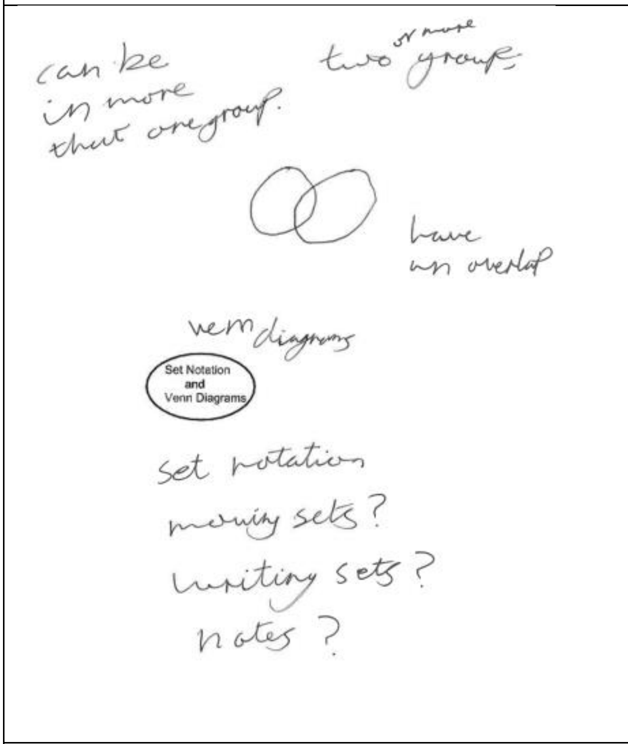
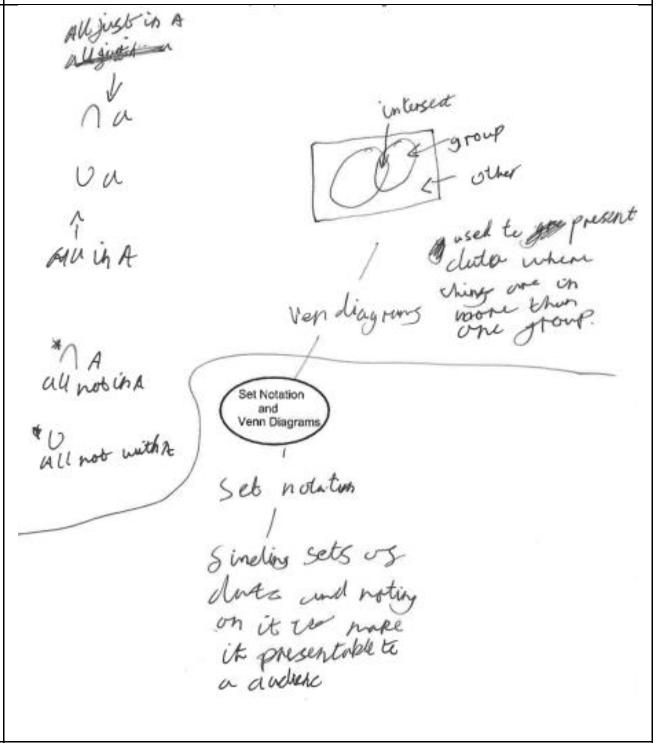
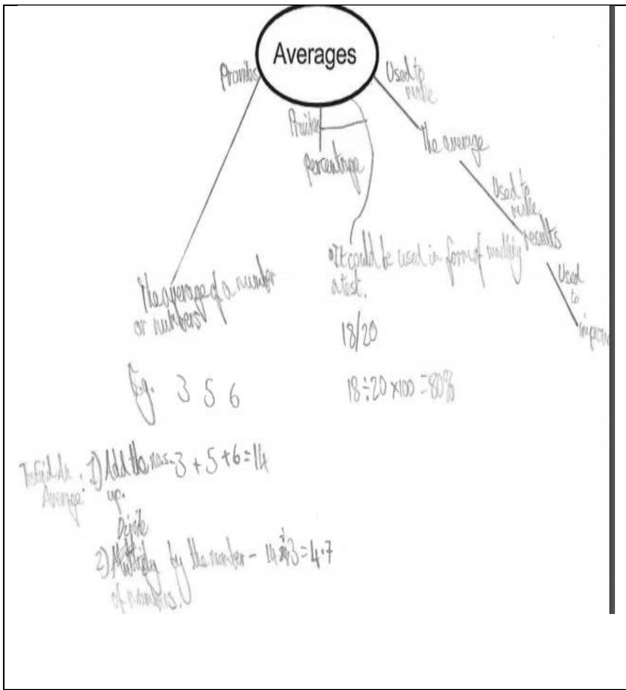
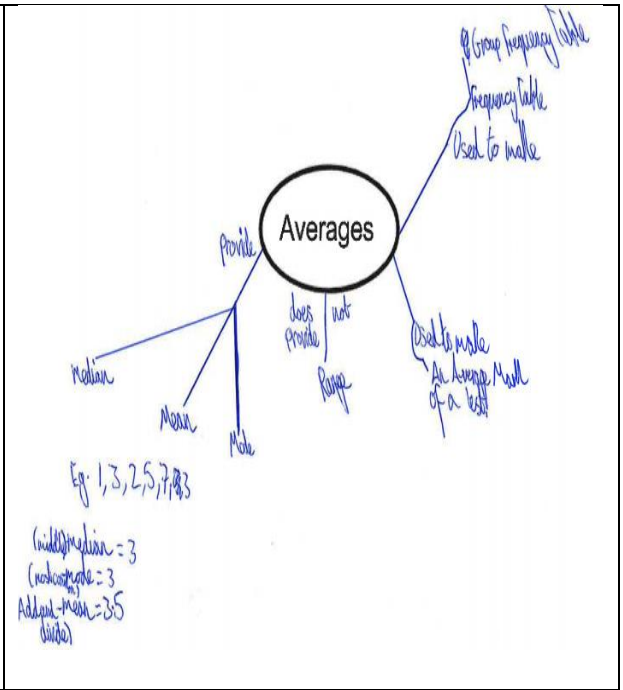
	
Cycle 2 pre-teaching with Frayer Model True score:1.44 Scaled score:58.1	Cycle 2 post-teaching with Frayer Model True score:3.35 Scaled score:73.06
	
Cycle 3 pre-teaching without Frayer Model True score:-1.13 Scaled score:35.68	Cycle 3 post-teaching without Frayer Model True score:1.35 Scaled score:56.69

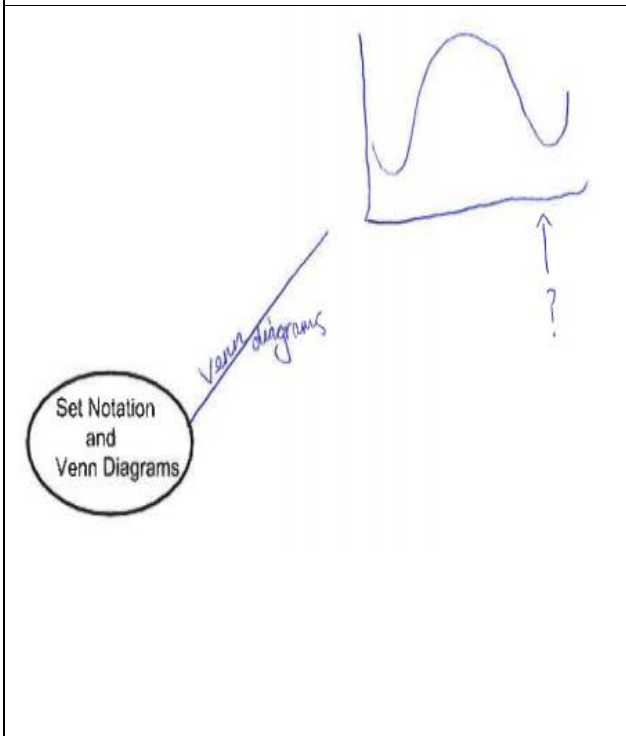
Figure 19. Shows the before and after scripts of Pupil C from Cycle 2 and 3



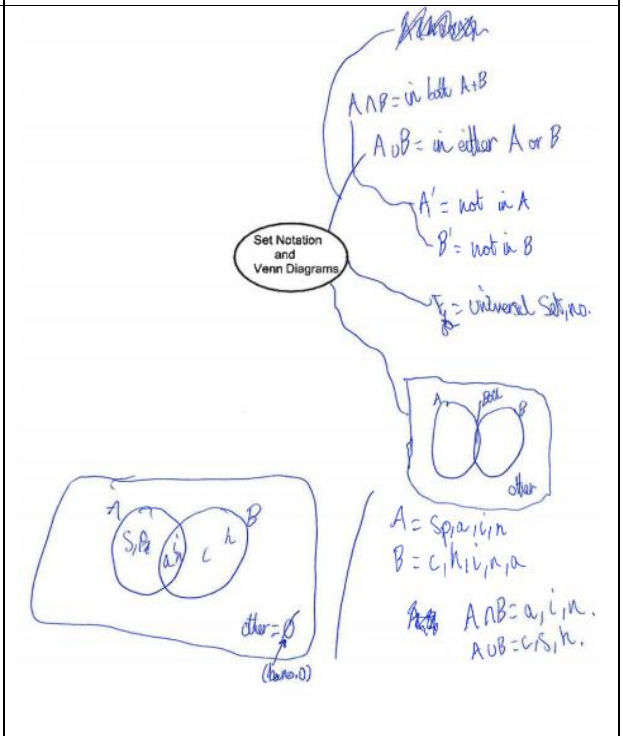
Cycle 2 pre-teaching with the Frayer Model
True score:-1.27
Scaled score:36.93



Cycle 2 post-teaching with the Frayer Model
True score:-2.52
Scaled score:27.13



Cycle 3 pre-teaching without the Frayer Model
True score:-4.36
Scaled score:8.35



Cycle 3 post-teaching without the Frayer Model
True score:2.33
Scaled score:64.97

Figure 20. Shows the before and after scripts of Pupil D from Cycle 2 and 3.

Reflect 3

The results of Cycle 3 will now be reviewed with respect to the two foci of this project.

1) To find a way of observing the development of conceptual understanding

Mind mapping has again proved itself to be effective in revealing pupils' misconceptions. Although Pupil D correctly described the union of two sets, their example indicated that they had not actually understood the concept. The issue of correctly memorised information being a misleading indicator of understanding is a problem raised by Vinner (2002). In this case the mind map was useful in identifying the misconception because their definition could be cross referenced with their example. However, it is possible that others could have concealed their misconceptions by not including examples and may have deceivingly scored higher which is an issue associated with the freedom of the mind mapping task.

Because of concerns about validity raised in Cycle 2, it was checked by calculating the Product Moment Correlation Coefficient (PPMCC) between pupils' CJ progress scores and MidYIS results. The MidYIS test is a baseline test conducted by all pupils at my school. It tests many skills, including mathematics, and supposedly indicates one's aptitude for learning (CEM 2018). The PPMCC between pupils' MidYIS results and their results from cycles 2 and 3 are r_2= -.43 and r_3= -.25, respectively. This shows there is a small negative correlation between the MisYIS results and data collected in this project and thus suggests that we are measuring something different to that of the MidYIS test. Although this suggests that the mind mapping method does not measure aptitude for learning it does not necessarily imply that it does not measure conceptual understanding as they are slightly different constructs.

Despite the review instructions given to pupils that encourage them to identify relationships between knowledge along the connecting lines, very few pupils actually did this. In the future, I would try this approach again after training pupils and familiarising them with mind mapping. This is because McClure et al. (1991) believes that a pupil's skill at mind mapping can influence the reliability and validity of the task. They held 90-minute workshops with pupils in their study prior to using and advocating them as an assessment task. Since it may be difficult to justify holding such workshops to stakeholders for the sole purpose of assessments, I would try to integrate mind mapping practise into lessons which would be easy to justify since they can also be used as a learning tool.

The CJ method in combination with the www.nomoremarking.com website has again been a positive experience. It proved useful as a platform for collaborative marking and was a fair and discrete way of accounting for colleagues' differing opinions. Because it allowed marking without a mark scheme, it accounted for the complexity of conceptual understanding but would also have reflected our learning goals for pupils. Since this study will help to determine whether use of the Frayer Model should be part of my typical practise, having the results reflect our school's learning goals is advantageous.

2) To observe how the Frayer Model impacts a pupil's understanding of a concept

The quantitative data indicates that both groups made advancements in their conceptual understanding over the teaching period. The experimental group made more progress, surpassing the control group on average by an additional 1/100 marks, and attaining a high calculated effect size. These encouraging results were also accompanied by a high significance level.

It was hoped that by looking for evidence of the features of the Frayer Model that its direct influence on a pupil's concept image may be observed. The figures in Table 10 indicate that pupils in both groups increased the number of definitions, facts/characteristics and examples they offered in the post-teaching assessment; although there was a greater improvement in the experimental group. This suggests that these features are developed through my typical teaching, but the Frayer Model is even more effective at producing these results. As aforementioned, these are also features of emerging conceptual understanding (Klausmeier 1992). It seems that focusing on these features of a concept in lessons makes this knowledge easier for pupils to recall and integrate into their existing understanding of the topic, resulting overall in a more comprehensive concept image.

Although the statistical calculations suggest that all pupils on average benefitted from taking part in the intervention, there was a wide variety of responses to the intervention. Both the pupils that made the most and least progress overall were from this group; their apparent developments over the intervention were strikingly different. The variety of responses to the intervention challenges the idea that the model is an easy template that will result in a more comprehensive concept image for all pupils. Although not in the scope of this project, it seems that further exploration would be useful to understand why pupils have responded so differently to the intervention. It could, for instance, be an influence of the secondary factors listed in the literature review.

The scripts of pupils that took part in both cycles were compared to see whether there were similarities in the way the model impacted their concept images. It seems that both post-intervention mind maps had a clearer structure than those created after typical teaching. Arguably, as insinuated by McClure et al. (1991), observing this difference is particularly impressive since by Cycle 3 the pupils were advantaged by having already encountered the assessment task. Regarding concept image, a structured mind map may indicate that the pupil has a clearer grasp of how knowledge about the concept is integrated. Adams (2010) too suggests this can be a consequence of using the model, believing that it helps pupils to connect and organise all the knowledge they have about a topic. However, because of flaws in the research design it is difficult to assume that these observed better structures can be attributed to the use of the Frayer Model.

For instance, the structure may be correlated to the level of prior knowledge known about the topic. Both pupils C and D had prior knowledge about averages and this coincides with the better structured mind maps. It seems possible that it may be easier to organise and integrate new knowledge into an already established and correct concept image. This is because it can be assimilated without making changes to its current structure (Piaget 1952). There were no errors in the pre-teaching scripts of both pupils, so the better structured mind maps may be a result of having prior knowledge rather than an influence of the Frayer Model.

In addition, the better structure could be attributed to the shorter period between teaching finishing and the assessment task being conducted in Cycle 2. The pupils in Cycle 3 waited one month, including a school holiday, before sitting their post-assessment as opposed to those in Cycle 2 that sat theirs the lesson after teaching had finished. Regarding Cycle 2, the short time period may have allowed pupils that do not have a deep understanding of the topic to present more than they have integrated into their concept image because it is easier for them to recall memorised information. Whereas those from Cycle 3 in a similar situation would find it harder to recall from memory. However, the time difference may not have affected the structure of the mind maps of those that did gain a deeper level of conceptual understanding, since Skemp (1976) and Coutinho (2014) believe this type of connected knowledge is easier to recall. That is, the short time period between the teaching and assessment task in Cycle 2 may have caused those that do not have a strong grasp of the concept to present better structured knowledge than they have really integrated into their concept image.

Although relationship dynamics may not directly impact the structure of the mind maps pupils offer, this factor may have influenced the apparent success of the intervention. Having never taught or encountered the pupils in Set 3 prior to the intervention, some lesson time was dedicated to learning names and establishing working relationships with pupils in the group. For these pupils, the intervention brought changes in the form of the Frayer Model as well as me as their teacher. This may have influenced their behaviour and thus interactions with the model by, for instance, making them more attentive or distracted. Whilst the pupils in Set 2, the control group, were meant to experience my typical teaching but may have still been influenced by the impact of encountering the model in Cycle 2. Repeatedly using Set 2 was not ideal as this may have reduced the apparent difference that using the model might make. This issue with the research design came about because of the nature of the action research model. Without a complete overview of the study from the start, decisions can be made that can unforeseeably be detrimental to subsequent cycles.

Conclusion

The Frayer Model presented a structured way to explicitly teach technical vocabulary without detracting from the focus of learning mathematics. Despite its simplicity and potentially large benefit to pupils, the model has not been given much direct attention in theoretical or empirical research, although it is often recommended in online professional literature. As an advocate of evidence-based practice, supportive research was required to justify integrating it into my teaching repertoire and recommending it to others. This evidence was gathered by using an action research project with Year 9 pupils at my school.

Responses on face value indicate that the model equipped pupils with definitions, facts/characteristics, examples and non-examples that they could use to communicate their understanding of a concept. However, if mind maps are assumed to be representative of a pupil's current understanding of a topic, then the results suggest that the model has promoted a more organised and comprehensive concept image than is typically achieved through my teaching. The theoretical literature explains that the model can do this in two ways. Firstly, the quadrants in the model create a focus on the limitations, context and negative knowledge of the concept (Klausmeier 1992; Watson & Mason 2002; Heemsoth & Heinze 2014). Secondly, by giving pupil's ownership to construct their own content, they have the opportunity to operate on and organise their knowledge of the concept and connecting ideas; which Piaget (1952) believes is required to develop conceptual understanding.

Implications

Wider Educational Community

This project has explored two concepts that require greater attention in the literature; the use of the comparative judgement method, since it is a relatively innovative approach to evaluating assessments, and the use of the Frayer Model in mathematics, because it seems to be an underrated teaching tool in this domain. As such, this study invites interest from both of these areas. However, due to the nature of practitioner research, compromises were made in this study that make the results very specific to my school and teaching environment and limit the appropriateness of generalising the conclusions further. Therefore, this project should only act as an incentive to the formal research community, inviting them to conduct more formal studies, and to other teachers wishing to conduct their own school-based studies.

Colleagues

Colleagues in my department collaborated on several occasions. They helped to select the vocabulary that was taught, mark the mind mapping task and two members were also teachers of the participating sets. The study may have impacted them personally in many ways, it may inspire them to try the model themselves or explore alternative methods of language development, or neither! But it is perhaps worth noting the wider potential implications this study could have for my school. Since the Frayer Model can be used across any domain, this could be an EAL strategy that is used universally at my school. It is possible that pupils may benefit from using it routinely in all lessons. As such, I will be presenting the Frayer Model to the whole school at their 'Marketplace' event in September, where teachers can share their resources.

Me as a researcher

This was the first practitioner-researcher based study I have conducted, and it was harder than expected. Researching in one's school limits the resources and time that can be allocated to it and expected of others, as well as the pool of pupils and colleagues that can be involved. I found it challenging to decide which compromises, that might decrease the validity of the study, should be made due to practicalities and ethical issues. However, there were more benefits than drawbacks of conducting my own study. By formally identifying the issue, it gave a clear focus and led me to search empirical and theoretical literature, which I would not typically refer to. This reminded me of the learning goals I have for my students and informed how tasks can be managed to promote conceptual understanding. Furthermore, trialling something new in the classroom and tracking its impact was exciting; I felt invested in the project, my pupil's results and my own professional development. I also feel that it gave me greater confidence with my teaching and that my approach to research methodology will improve if I continue conducting my own studies. Although I would advocate practitioner research to other teachers, I think that they should ensure they have a support network in place where tough decisions about compromises can be discussed.

Me as a teacher

This study suggests that integrating the Frayer Model into my regular teaching practice will benefit the conceptual development of my pupils beyond my current practice. Because it is so easy to implement and seems to be a manageable and sustainable change, I intend on making it part of my teaching repertoire. It will not be used for every technical word associated with a topic like in this study, to avoid disengaging pupils with its overuse. I may also occasionally alter the structure of the model, since not all quadrants seemed suitable for every word. Furthermore, I intend on trialling it with all my classes. Since the task seems differentiated when pupils make their own content, I cannot see why the conceptual understanding of pupils of all ages would not benefit from engaging with the Frayer Model.

Appendices

Appendix 1 Klausmeier's (1992) level descriptors

‘A concept has been formed at the concrete level when the learner recognizes an item as the same one previously encountered when that item is encountered in the identical spatial orientation or other context in which it was initially encountered. The mental processes involved in forming a concept at the concrete level are attending to an item (a drawing of an equilateral triangle on a page of a book), discriminating the item as an entity different from its surroundings (a drawing of a right triangle and an isosceles triangle) on the basis of one or more of its defining attributes (shape), representing the item in long-term memory as a visual image, attending to the item when it is again encountered in the identical context (on the same page and the learner seated in the same position), retrieving the representation (image), and using it in recognizing the item as the same entity discriminated earlier, The learner does not simply discriminate the item but also recognizes it as an entity having permanence.

A learner demonstrates formation at the identity level by recognizing an item (a drawing of an equilateral triangle) as the same one previously encountered when the item is observed from a different spatiotemporal perspective (in a different position or a different orientation on the page) or when it is sensed in a different modality (the national anthem when heard sung and when seen as a music score). The mental processes at the identity level include those at the concrete level and, in addition, a new process: generalizing that the item, though experienced differently, is the same one discriminated earlier.

Formation at the classificatory level requires recognizing at least two items as being equivalent. In forming the classificatory level of any given concept (equilateral triangle), the learner carries out the processes at the identity level and a new process: generalizing that at least two different entities are equivalent (a smaller and a larger equilateral triangle). Formation at the classificatory level continues until the learner can identify all or nearly all examples and can discriminate nonexamples (right triangle and isosceles triangle) but still cannot accurately explain the basis for the classification (length of sides, equality of angles). Before considering the formal level, be aware that having the name of the concept is not prerequisite for forming a concept at the concrete, identity, or classificatory level; however, it facilitates all three.

A learner demonstrates the formal level (equilateral triangle) by being able to identify any and all examples and nonexamples (other kinds of triangles and other geometric forms), name the concept and its defining attributes (three equal sides, three equal angles), give the experts' definition of the word(s) that name(s) the concept (a plane, simple, closed figure of three equal sides and three equal angles), and specify the critical attributes (three equal sides, three equal angles) that differentiate the concept from other closely related concepts (right triangles, isosceles triangles). To achieve the formal level the learner must be able to carry out the processes of the classificatory level and either a set of inductive processes that includes hypothesizing, evaluating, and inferring or a set of meaningful reception operations that includes processing and correctly using the given information.’

(Klausmeier 1992, pp.272–273)

Appendix 2 Time line of practitioner research

Week Beginning	Description
23/10/17	• Email to headmaster requesting permission to conduct research sent.
30/10/17	• Permission to conduct research received from headmaster.
05/11/17	• Email to CUREC requesting permission to conduct study sent.
11/12/17	• Permission from CUREC to proceed with study received.
08/01/18	• Cycle 1 (pilot) with Year 7 conducted and uploaded to www.nomoremarking.com • Colleagues invited to take part in vocabulary selection process.
15/01/18	• Vocabulary selection for Cycle 2 takes place in department meeting.
22/01/18	• Pre-teaching assessment conducted with Year 9 classes and teaching of averages begun.
29/01/18	• Post-teaching assessment conducted with Year 9 classes.
05/02/18	• Comparative judgement of cycle 2 scripts in department meeting.
12/02/18	• Email to Deputy Head Academic requesting permission to swap classes sent and received.
19/02/18	• Vocabulary selection for Cycle 3 takes place in department meeting. • Email requesting volunteers for teacher swap sent and offers received.
12/03/18	• Pre-teaching assessment of venn diagrams conducted with Year 9 Set 2.
19/03/18	• Pre-teaching assessment conducted with Year 9 Set 3 and teaching of Venn Diagrams and Set Notation.
26/03/18	• Teaching of Venn Diagrams and Set Notation to year 9 set 2.
23/04/18	• Post-teaching assessment conducted with year 9 set 3.
30/04/18	• Post-teaching assessment conducted with Year 9 Set 2.
21/05/18	• Comparative judgement of Cycle 3 scripts in department meeting.

Appendix 3 CUREC Approval Letter

Dear [redacted],

How does the Frayer Model impact pupils' concept image formation in mathematics?

The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants. Please note however that data collection should not have started until this approval is given.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

If your research involves participants whose ability to give free and informed consent is in question (this includes those under 18 and vulnerable adults), then it is advisable to read the following NSPCC professional reporting requirements for cases of suspected abuse

http://www.nspcc.org.uk/globalassets/documents/information-service/factsheet-child-abuse-reporting-requirements-professionals.pdf

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to [redacted] for consideration.

Good luck with your research study.

Sincerely

[redacted]
Chair
Department of Education
University of Oxford

[redacted]

Senior Research Fellow
Harris Manchester College
University of Oxford
[redacted]

Associate Professor
Department of Education
University of Oxford

Appendix 4 Email to the Headmaster

UNIVERSITY OF OXFORD

[redacted]

30/10/17

Dear [redacted],

I am writing to obtain consent to conduct research in your school this year as part of my Masters in Learning and Teaching. My project is titled 'How does the Frayer Model impact pupils' concept image formation in mathematics?' and is supervised by [redacted] It aims to improve pupils' understanding of mathematical concepts via a vocabulary intervention.

The intervention will take place with my Year 9 mathematics set. My Year 7 class and another Year 9 class will be used for the pilot study and control group, respectively. The process will also require the input of my colleagues in the Mathematics Department. I will request their opinion on pupil work and key vocabulary in the curriculum.

Oxford University has strict ethical procedures on conducting research with teachers and young people, consistent with current British Educational Research Association guidelines. As my research project is classified as normal working practice, your consent is sufficient on behalf of pupils and their parents. Teachers will be provided the opportunity to opt out of collaborating on the project.

All participants and the school will remain anonymous in the report and all results will be collected by the Easter holidays. If you need more information about what is involved, please contact me. I would be grateful if you would complete the form below and return it via email or hard copy to my pigeon hole.

Thank you for your time and attention. I look forward to hearing from you.

Regards,

[redacted]

[redacted]

[redacted]

☐We do not wish to participate in this project.

☐We would like to take part in this project.

Please return this form via email.

Appendix 5 Email asking for volunteers to participate in cycle 2

Dear all,

As part of my master's project, I will be conducting an intervention with my Year 9 class. The intervention will last for the duration of the teaching of Averages. In order to compare the progress of my class with another group, I would like another class to take an assessment before and after the teaching of this topic. The assessment is straightforward, as pupils have to create a mind map where they write down everything they know about Averages. It must be done in exam conditions and they must be allowed as much time as they need to complete their mind map. I'll give further details if you decide to volunteer your group.
Please email me if you have any further questions or are willing to volunteer your class in the study. Your name will be anonymised in the final assignment.

Kind regards,

[redacted]

Appendix 6 Email regarding vocabulary selection process

Dear all,

As most of you are aware, I will be conducting research at the school this year as part of studying for my masters at the University of Oxford. The title of my study is 'How does the Frayer Model impact concept image formation of Year 9 pupils in mathematics?'. The model uses key vocabulary to develop understanding.

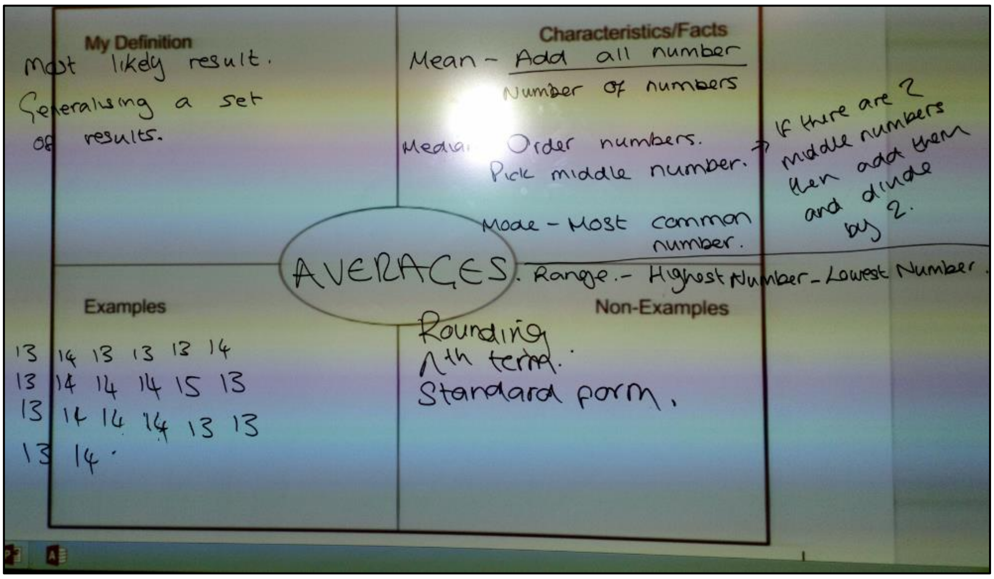
To make sure that I don't miss out any key terms, I would like to note down the key vocabulary we believe our Year 9 pupils should know about Averages. The words identified will then be used with the model in lessons. Your individual contributions will be unidentifiable and you will be kept anonymous. I hope to do this in the last 15 minutes of next week's department meeting.

Please contact me via email or in person if you have any questions about this activity or my study. If you do not wish to participate, you will be given the option of leaving before the activity begins.

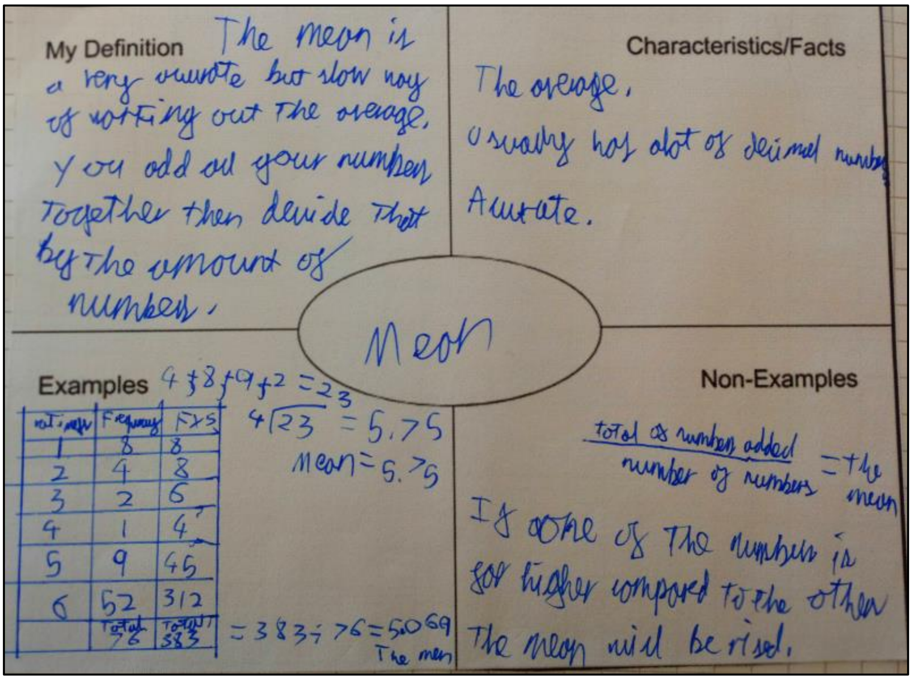
Kind regards,

[redacted]

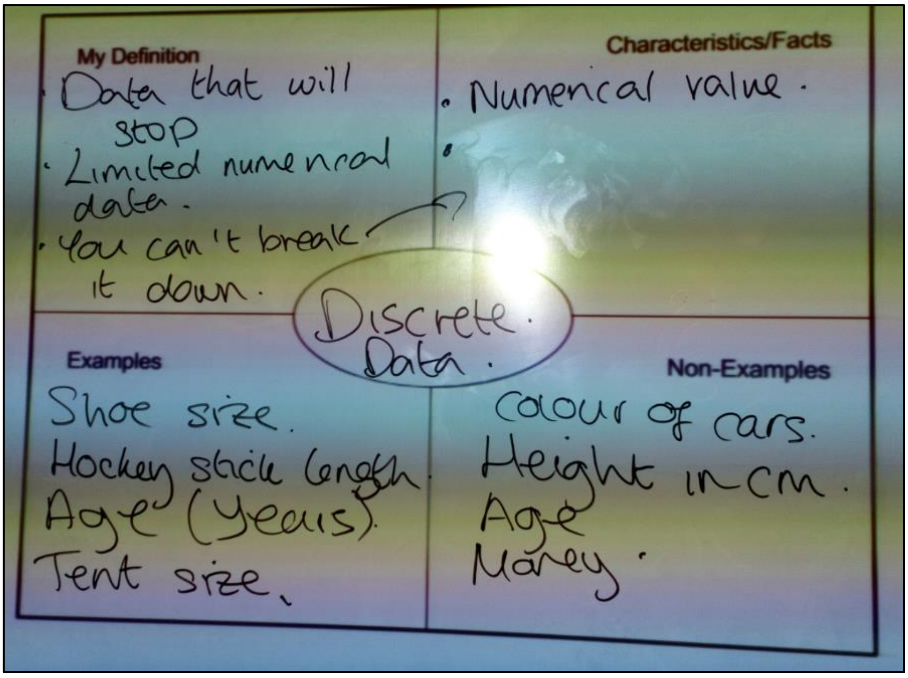
Appendix 7 Examples of how the Frayer Model was used in Cycle 2



This shows a photograph of the projected Frayer Model that was completed using contributions from the class.



This is a photograph of a pupil's reflective Frayer Model after teaching about the mean.



This is a photograph of the projected Frayer Model used to introduce the concept of Discrete Data and was completed collectively by the pupils and me.

Appendix 8 Email to the Deputy Head, seeking permission to swap classes for a week

Dear [redacted],

I'm writing to seek permission to do a 'teacher swap' within Year 9. Last term I tried out my intervention with my Year 9 class. The results compared my class with the progress made by set 1 and they were really promising. In order to isolate factors to determine whether the intervention caused this result, my supervisor has asked whether it is possible for me to use this intervention again with another group in Year 9. This would involve me swapping with another teacher for one week. I will ask the department for volunteers to take part. The approved project title is about the use of the intervention with Year 9 pupils, which is why I need to continue with this year group. The intervention and collection of results should be completed by the Easter holidays.

Although swapping can be a pain in terms of not being familiar with the class, I think that it could be a beneficial project for the other teacher involved to do and would not cause harm to the progress of the pupils. Please can I have permission to go ahead with this?

Regards,

[redacted]

Appendix 9 Email asking for volunteers to participate in Round 3

Dear all,

I am looking for volunteers to participate in the next phase of research at school. This would involve swapping classes with a Year 9 teacher for a week at the end of term, so that I would teach your Year 9 group and you would teach mine. During this week, I will teach your group about Venn Diagrams and Set Notation whilst you teach mine about Shapes from the scheme of work. Before and after teaching your group, I will conduct a short assessment to measure the progress they make. Your name will be anonymised in the final assignment. Permission to do this has already been authorised by [redacted].

I know that swapping classes, especially if you don't know the pupils, can be stressful. However, I'm hoping that you may realise the benefits of participating as well. Please email me if you are interested in taking part or have any further questions.

Kind regards,

[redacted]

References

Abi-El-Mona, I. & Adb-El-Khalick, F., 2008. The Influence of Mind Mapping on Eighth Graders' Science Achievement. *School Science and Mathematics*, 108(7), pp.298–312. Available at: http://doi.wiley.com/10.1111/j.1949-8594.2008.tb17843.x.

Adams, A., 2010. Rehearsal or Reorganisation: Two Patterns of Literacy Strategy Use in Secondary Mathematics Classes. *The Mathematics Enthusiast*, 7(2), pp.371–390.

Adler, J., 1999. The Dilemma of Transparency: Seeing and Seeing through Talk in the Mathematics Classroom. *Journal for Research in Mathematics Education*, 30(1), pp.47–64.

Babione, C. & Anderkin, C., 2015. *Practitioner Teacher Inquiry and Research*, San Francisco: John Wiley & Sons.

Barwell, R., 2011. Some issues concerning EAL in the mathematics classroom. *NALDIC*. Available at: https://www.naldic.org.uk/Resources/NALDIC/Initial Teacher Education/Documents/Maths1.pdf [Accessed December 23, 2016].

BERA, 2018. *Ethical Guidelines for Educational Research*, London. Available at: https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018.

Bills, L. et al., 2006. Exemplification in Mathematics Education. In J. Novotna et al., eds. *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education*. Prague, pp. 126–154.

Bills, L. & Rowland, T., 1999. Examples, Generalisation and Proof. *Research in Mathematics Education*, 1(1), pp.103–116.

Bisson, M.-J. et al., 2016. Measuring Conceptual Understanding Using Comparative Judgement. *International Journal of Research in Undergraduate Mathematics Education*, 2(2), pp.141–164.

Carpenter, T. & Lehrer, R., 1999. Teaching and Learning Mathematics with Understanding. In E. Fennema & T. Romber, eds. *Mathematics Classrooms that Promote Understanding*. University of Wisconsin-Madison: Routledge.

CEM, 2018. THE CENTRE FOR EVALUATION AND MONITORING AT DURHAM UNIVERSITY. *Centre for Evaluation & Monitoring, Durham University*. Available at: https://www.cem.org/about-us [Accessed July 17, 2018].

Clarkson, P.C., 1992. Language and Mathematics: A Comparison of Bilingual and Monolingual Students of Mathematics. *Educational Studies in Mathematics*, 23(4), pp.417–429.

Cline, T. & Frederickson, N., 1996. *Curriculum related assessment, Cummins and bilingual children*, Clevedon : Multilingual Matters .

Cockcroft, W.H., 1982. *Mathematics Counts: Report of the Committee of Inquiry into the Teaching of Mathematics in Schools*, London.

Cohen, M. & Carpenter, J., 1980. The effects of non-examples in geometrical concept acquisition. *International Journal of Mathematical Educational in Science and Technology*, 11(2), pp.259–263.

Concepcion, M., 2012. *The problem with math is English : A language-focused approach to helping all students develop a deeper understanding of mathematics* 1st ed., San Francisco: Jossey-Bass.

Coutinho, E., 2014. Concept Maps: Evaluation Models for Educators. *Journal of Business and Management Sciences*, 2(5), pp.111–117.

Cummins, J., 2000. *Language, power and pedagogy: bilingual children in the crossfire*, Clevedon: Multilingual Matters.

Cummins, J., 1976. The Influence of Bilingualism on Cognitive Growth: A Synthesis of Research Findings and Explanatory Hypotheses. *Bilingual Education Project*.

Dawe, L., 1983. Bilingualism and mathematical reasoning in English as a second language. *Educational Studies in Mathematics*, 14(4), pp.325–353.

Dunston, P. & Tyminski, A., 2013. *What's the big deal about vocabulary?*,

Edwards, A. & Talbot, R., 1994. *The hard-pressed researcher A research handbook for the caring professions*, Longman Group Limited.

EU, 2016. General Data Protection Regulation. *Official Journal of the European Union*. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN.

Evrekli, E., Inel, D. & Balim, A.G., 2010. Development of a scoring system to assess mind maps. *Procedia - Social and Behavioral Sciences*, 2(2), pp.2330–2334.

Feldman, A. et al., 2007. *Teachers Investigate Their Work, An Introduction to Action Research across the Professions*, Routledge Taylor & Francis Group.

Frayer, D., Fredrick, W. & Klausmeier, H., 1969. *A Schema for Testing the Level of Concept Mastery*, Madison.

Freeman, L., Urbaczewski, A. & Croasdell, D., 2003. Concept Maps for Teaching and Assessment. *Communications of the Association for Information Systems*, 12(396–405).

Fuchs, L. et al., 2017. Fraction Intervention for Students with Mathematics Difficulties: Lessons Learned from Five Randomized Controlled Trials. *Journal of Learning Disabilities*, 50(6), pp.631–639.

Heemsoth, T. & Heinze, A., 2014. The impact of incorrect examples on learning fractions: A field experiment with 6th grade students. *Instructional Science*, 42(4), pp.639–657.

Issa, T. & Eve, P., 2011. Making Maths curriculum more accessible: Strategies for children learning EAL. *NALDIC*. Available at: https://www.naldic.org.uk/Resources/NALDIC/Initial Teacher Education/Documents/MathematicsPrimary.pdf.

Jennings, D., 2012. *The Use of Concept Maps for Assessment*. UCD Dublin. Available at: https://www.ucd.ie/t4cms/UCDTLA0040.pdf.

Jones, I. & Inglis, M., 2015. The Problem of Assessing Problem Solving: can Comparative Judgement Help? *Educational Studies in Mathematics*, 89(3), pp.337–355.

Jones, I., Swan, M. & Pollitt, A., 2014. Assessing Mathematical Problem Solving Using Comparative Judgement. *International Journal of Science and Mathematics Education*, 13(1), pp.151–177.

Kilpatrick, J., 2001. Understanding Mathematical Literacy: The Contribution of Research. *Educational Studies in Mathematics*, 47(1), pp.101–116.

Klausmeier, H., 1992. Concept Learning and Concept Teaching. *Educational Psychologist*, 27(3), pp.267–286.

Klausmeier, H. & Frayer, D., 1970. *Cognitive Operations in Concept Learning*, Madison.

Lave, J. & Wenger, E., 1991. *Situated Learning: Legitimate Peripheral Participation*, Cambridge University Press.

Lefevre, P. & Hiebert, J., 1986. *Conceptual and Procedural Knowledge: The Case of Mathematics* J. Hiebert, ed., Routledge.

Mariotti, M.A. & Fischbein, E., 1997. Defining in Classroom Activities. *Educational Studies in Mathematics*, 34, pp.219–248.

Mason, J. & Watson, A., 2005. *Mathematics as a Constructive Activity: Learners Generating Examples*, New Jersey: Lawrence Erlbaum Associates.

McClure, J., Sonak, B. & Suen, H., 1991. Concept Map Assessment of Classroom Learning: Reliability, Validity, and Logistical Practicality. *Journal of Research in Science Teaching1*, 36(4), pp.475–492.

Monaghan, F., 2009. Mapping the Mathematical Langscape. In R. Barwell, ed. *Multilingualism in Mathematics Classrooms Global Perspectives*. Bristol: Multilingual Matters.

Monroe, E. & Pendergrass, M., 1997. *Effects of Mathematical Vocabulary Instruction on Fourth Grade Students*,

NMM, 2018. No More Marking Blog. Available at: https://blog.nomoremarking.com/ [Accessed July 31, 2018].

Novak, J.D., 1990. Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching*, 27(10), pp.937–949.

Pelton, R., 2010. *Action Research for Teacher Candidates: Using Classroom Data to Enhance Instruction*, R&L Education.

Peters, C., 1975. The Effect of Systematic Restructuring of Material upon the Comprehension Process. *International Literacy Association*, 11(1), pp.87–111.

Piaget, J., 1952. *The Origins of Intelligence in Children* M. Cook, ed., New York: International Universities Press, Inc.

Powney, J. & Watts, M., 1987. *Interviewing in Educational Research*, London: Routledge.

Price, C. & van Jaarsveld, P., 2017. Using Open-response Tasks to Reveal the Conceptual Understanding of Learners-Learners Teaching the Teacher what they Know about Trigonometry. *African Journal of Research in Mathematics*, 21(2), pp.159–175.

Reeves, J., Redford, M. & McQueen, I., 2010. Practitioner research and excellence in teaching. *Scottish Educational Review*, 42(2), pp.74–88.

Riccomini, P. et al., 2015. The Language of Mathematics: The Importance of Teaching and Learning Mathematical Vocabulary. *Reading and Writing Quarterly*, 31(3), pp.235–252.

Rittle-Johnson, B. & Alibali, M.W., 1999. Conceptual and Procedural Knowledge of Mathematics: Does One Lead to the Other? *Journal of Educational Psychology1*, 91(1), pp.175–189.

Short, D., 1993. Assessing Integrated Language and Content Instruction. *TESOL Quarterly*, 27(4), pp.627–656.

Skemp, R., 1976. Relational Understanding and Instrumental Understanding. *Mathematics teaching*, 77, pp.20–26.

Smith, L., 1993. *Assessment of Student Learning in Mathematics*, Columbia.

Star, J., 2005. Reconceptualising Procedural Knowledge. *Journal for Research in Mathematics Education*, 36(5), pp.404–411.

Tall, D. & Vinner, S., 1981. Concept Image and Concept Definition in Mathematics with Particular Reference to Limits and Continuity. *Educational Studies in Mathematics*, 12(2), pp.151–169.

Thurstone, L., 1994. A Law of Comparative Judgement. *Psychological Review*, 101(2), pp.266–270.

Turkay, S., Concept Maps Are They Good For Assessment? Available at: https://hilt.harvard.edu/files/hilt/files/concept_maps_as_assessement_tools.pdf [Accessed January 3, 2018].

UNICEF, 2010. *A summary of the UN convention on the rights of the child*, UK. Available at: https://downloads.unicef.org.uk/wp-content/uploads/2010/05/UNCRC_summary-1.pdf?_ga=2.95439188.972467074.1531741127-1087862866.1531741127.

Vinner, S., 1991. The Role of Definitions in the Teaching and Learning of Mathematics. In *Advanced Mathematical Thinking*.

Vygotsky, L., 1986. *Thought and Language* A. Kozulin, ed., Cambridge: MIT Press.

Wanjiru, B. & O-Connor, M., 2015. Effects of Mathematical Vocabulary Instruction on Students' Achievement in Mathematics in Secondary Schools of Murang'a County, Kenya. *Journal of Education and Practice*, 6(18), pp.201–207.

Warren, E. & Cooper, T., 2009. Developing Mathematics Understanding and Abstraction: The case of the Equivalence in the Elementary Years. *Mathematics Education Research Journal2*, 21(2), pp.76–95.

Watson, A., 2006. *Raising achievement in secondary mathematics*, Maidenhead: Open University Press.

Watson, A. & Mason, J., 2002. Student-Generated Examples in the Learning of Mathematics. *Canadian Journal of Math, Science & Technology Education*, 2(2), pp.237–249.

Webb, L. & Webb, P., 2008. Introducing Discussion into Multilingual Mathematics Classrooms: An Issue of Code Switching? *Pythagoras*, (67), pp.26–32.

Wilson, E., 2017. *School-based Research A guide for education students* E. Wilson, ed., SAGE.

Zandieh, M. & Rasmussen, C., 2010. Defining as a Mathematical Activity: A Framework for Characterising Progress from Informal to more Formal ways of Reasoning. *The Journal of Mathematical Behaviour*, 29, pp.57–75.

Zaslavsky, O. & Shir, K., 2005. Students' Conceptions of a Mathematical Definition. *Journal for Research in Mathematics Education*, 36(4), pp.317–346.