Using interleaved practice in a mathematics secondary classroom: effects and perceptions

Ruth Dodson

A Research & Development Project

Submitted for the MSc Learning & Teaching 2020

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	Dodson
First Name	Ruth
Faculty Board	Education
Title of Dissertation	Using interleaved practice in a mathematics secondary classroom: effects and perceptions

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.

Table of Contents

1. Introduction

Many mathematics teachers will recognise this format of structure in their lessons: starter of mixed questions, new skill is taught, new strategy is practised by pupils, finished with some form of plenary - usually based on the same concept. It is common in classrooms to give learners plenty of practice on the skill they have been taught. Given that so much time during a lesson is allocated to practice of those skills, it seems important to establish the most effective way of doing so.

It is this structure of practice which forms the focus of this research project. The standard presentation of skill, discussed above, whereby practice of a strategy or skill is grouped together, is an example of blocked practice (e.g. Rohrer, Dedrick & Burgess, 2014; Foster, Mueller, Was, Rawson, Dunlosky, 2019). Whilst this process may be beneficial for short term retention, cramming before an exam for example, it is thought that mixing problems and strategies up during practice will be more beneficial long term (e.g. Mayfield & Chase, 2002; Taylor & Rohrer, 2010; Rohrer et al., 2014; Foster et al., 2019). When practice is structured in such a way, with the strategy required for a problem presented in a non-consecutive manner, it is known as interleaved practice.

Interleaved practice has been proven to be a superior method of practice across a range of both non-educational and educational settings (e.g. Kang & Pashler, 2012; Rohrer et al., 2014). In those educational settings, and specifically in mathematics, it has been found that interleaving improved test results, strategy association and confidence (e.g. Mayfield & Chase, 2002; Taylor & Rohrer, 2010; Rohrer et al., 2014; Foster et al., 2019).

My current school setting is an independent coeducational secondary school in Oxfordshire, where my role is Deputy Head of Mathematics. Over the last two years the department have been focusing on introducing mixed topic homework and low stakes quizzes. Both of which contained problems whereby the strategies required are mixed up (interleaved). In discussions with pupils, I discovered that their concern was the juxtaposition between how they found questions in the lesson compared to these homework tasks and quizzes. Similarly, a key motivator for this project was the pupils' lack of

confidence in their end of year assessment. When asked, pupils seemed to be similar in their response: "when we do it in the lesson, I can do it, but I didn't know in the examination to do that". Teachers will recognise the familiar sound of "Oh it's that, I can do that!" when going through examination questions with a group. When the pupil is told the topic or strategy, they can execute it, but they struggle to identify the strategy required in the first place. This formed my desire to question why this was; why do students who perceive to understand during lessons, seem unable to associate that strategy to its relevant problem? I wanted to investigate ways in which the department can improve this. I found myself wondering whether we were providing enough opportunity for pupils to associate solutions to problems, or whether we were teaching them in a way which removed their need to identify strategies.

Having explored the research pertaining to the positive effects interleaved practice can have on test results in mathematics (e.g. Mayfield & Chase, 2002; Rohrer & Taylor, 2010; Rohrer, Dedrick & Hartwig, 2020a; Foster et al., 2019), the aim of the study was to ascertain whether this effect can be experienced with pupils in my setting. On a similar note, having recognised a gap in the research with respect to pupil perceptions of their experiences of interleaved practice, this became an important parallel focus for the study.

Therefore, the intervention in this research introduced interleaved practice during lessons with my teaching class. Measuring the outcome of such intervention with a test, questionnaire, and interview. During the process of this intervention, COVID-19 pandemic resulted in the closure of schools for the last 13 weeks of the academic year and, therefore, changes to the study had to be made. This will be discussed in more detail later. As the school closed and lessons moved online, in response to the positive effects found from the initial data collection (test, survey and interview), the study was adapted to explore how pupils perceive interleaved practice could be used in remote lessons.

The research questions posed in the finalised study were:

1. What effects did the use of interleaved practice in lessons have on test results?

2. What were pupil perceptions of interleaved practice in lessons?
3. How can interleaved practice be used in remote learning?

2. Literature Review

The literature review begins by exploring the definitions associated with interleaved practice, followed by an exploration of previous studies which have implemented interleaved practice, in various contexts. Concurrently, it is recognised that interleaved practice is not an isolated learning strategy; if interleaved practice is adopted then it will follow that spaced practice and retrieval practice is also taking place, which are both recognised and well researched phenomena (e.g. Yazdani & Zebrowski, 2006; Pyc & Rawson, 2007; 2009; Rohrer et al., 2014; Rohrer, Dedrick, Hartwig & Cheung, 2020b). As both retrieval practice (the expression of knowledge to an applicable level), and spaced practice (spacing out the presentation of problems), are inherent benefits of interleaved practice it is important to also acknowledge and define these in the literature review.

As this study will measure the impact of introducing interleaved practice to lessons, current literature exploring the use of interleaved practice to improve test results will be discussed, addressing how the timing of testing and intervention design produces different results (e.g. Rohrer et al., 2015; Nazari & Ebersbach, 2019). Whilst it is important to explore attainment impacts of interleaving, it is equally important, if not more important, to address what the experience of interleaved practice implementation is like for the participants. Therefore, the literature review will close on exploring the feasibility for teachers with the restricting resources available (e.g. Yazdani & Zebrowski, 2006; Rohrer & Taylor, 2007; Casey, Carroll & Crowley, 2018), followed by the desirable difficulties (e.g. Taylor & Rohrer 2010; Rohrer et al., 2020b) and the increased engagement (e.g. Metcalfe & Xu, 2016) experienced by the learners.

2.1 What defines Interleaved Practice?

2.1.1 Interleaved practice vs. blocked practice

A significant factor in the process of learning within a mathematics lesson is the practice of taught skills. The structure and focus of that practice is at the foundation of interleaved practice. Interleaved practice involves pupils working through different types of problems during that practice, be that different skills in superficially similar questions or entirely different skills for dissimilar problems as defined here by Foster et al. (2019):

> "Interleaved practice involves studying exemplars from different categories in a non-systematic, pseudorandom order under the constraint that no two exemplars from the same category are presented consecutively" (p. 1088)

The non-consecutive presentation of strategy requirements is a definitive factor of interleaved practice that is well recognised. The nature of that strategy is further defined by Rohrer et al. (2020b):

> “problems within an assignment are arranged so that no two consecutive problems require the same strategy, where strategy is defined loosely to include a procedure, formula, or concept.” (p. 40)

The imperative for both definitions is the notion that strategies should be intermixed, and a learner should not be able to continue to solve problems using the previously deployed strategy.

This is contrasted by blocked practice. In blocked practice pupils are taught a topic and practise that topic before moving on to the next topic. Rohrer et al. (2014) define blocked practice as the "grouping of problems by strategies” (p. 1323).

If a learner were faced with strategy a, b, c, d then the presentation of those strategies would look very different in interleaved practice compared to blocked practice:

Interleaved Practice:	abcd abcd abcd abcd	or	abcd acdb cabd bdca
Blocked Practice:	aaaa bbbb cccc dddd		

Whilst interleaved practice may present itself during revision lessons or mixed starters, the focus of practice in most mathematics lessons is on the topic and strategy being taught. The blocked practice

of concentrated new skills tends to be how mathematics lessons are largely taught and in fact are how most resources, including textbooks, are structured (Rohrer et al., 2020a).

There may be some benefits to being presented with a topic in a blocked manner, such as overlearning. Overlearning is the deliberate overtraining of a task, with students practising a concept beyond the point where they have learnt the skill (Rohrer & Taylor, 2006; 2007). Overlearning has been found to have positive effects on short term retention and immediate test performance (Driskell et al., 1992). Driskell et al. (1992) found, in their meta-analysis of 15 studies on overlearning, that there was a moderate improvement in retention from overlearning, although this was predominantly found in cognitive tasks. However, positive effects from overlearning have not always been found. Rohrer and Taylor (2006) found that overlearning produced no observable benefit to student retention (tested at both one week and four weeks). There is also very limited research revealing the long-term effects of overlearning in mathematics.

Notably, more recent literature (Rohrer et al., 2006; 2007; Rohrer et al., 2020b) contradicts those findings, of the positives of blocked practice (and overlearning), in favour of interleaved practice. Alongside claims that blocked practice reduces the difficulty of a mathematics problem (Rohrer, Dedrick & Stershic, 2015). Many studies provide evidence in support of using interleaved practice within the learning stage, rather than just as a revision or assessment tool, these such studies should be addressed in more detail.

2.1.2 Studies exploring interleaved practice compared to blocked practice

Many studies compare the efficacy of interleaved vs. blocked practice, finding that introducing or increasing the amount of interleaved practice produced benefits. These benefits have been explored over a variety of cognitive tasks and content. Earlier investigation of interleaving began with motor skills (Bjork, 1994), progressing into category learning paradigms (Kornell & Bjork, 2008). An example of this, was interleaving the presentation of artists which formed the basis of Kornell and Bjork's (2008) category learning paradigm, comparing blocked vs. interleaved learning. They measured a greater

recognition of novel paintings by those participants of the interleaved presentation group. Their study design was also replicated later by Kang and Pashler (2012) with the same findings.

Whilst much of the earlier research on interleaved practice, as mentioned above, has demonstrated the positive effects of interleaved practice in non-educational settings, the research has recently been extended into an educational focus. As the purpose of this study is to explore interleaved practice within my own setting, as a mathematics teacher, it is important to narrow and focus the exploration of this concept within the mathematical domain.

Many randomised controlled studies have been carried out to measure interleaving effects over a wide range of mathematical concepts, in fact Dunlosky et al. (2013) found that the studies were most promising in mathematics. Specifically, in mathematics, benefits of interleaving have been largely demonstrated in laboratory settings (e.g. Mayfield & Chase, 2002; Rohrer & Taylor, 2007; Le Blanc & Simon, 2008; Taylor & Rohrer, 2010; Foster et al., 2019). However there is also a plethora of classroom studies, covering a broad spectrum of interleaving in mathematics: Rau, Aleven & Rummel (2010; 2013) –fractions, Ziegler & Stern (2014) – addition and multiplication problems, Rohrer Dedrick & Burgess (2014); Rohrer Dedrick and Stershic (2015); Ostrow, Heffernan, Heffernan & Peterson (2015) - geometry.

One pivotal study is that by Rohrer & Taylor (2007), which has been replicated numerous times since (e.g. Le Blanc & Simon, 2008; Foster et al., 2019). In their research, college students were taught how to find the volume of the four solids, followed by two randomly assigned practice schedules. One with blocked practice, one with interleaved (which they called mixed in the study). The mixed practice produced superior results to the blocked group on the test, one week later. A finding which was further supported with Le Blanc & Simon's (2008) study which replicated the volume problems used by Rohrer & Taylor (2007).

In yet another laboratory-based study based on Rohrer and Taylor (2007)'s design, Foster et al. (2019) examined the level of interleaved practice to determine why the effects of interleaving work. The four

sample groups were randomly assigned to a group; blocked, interleaved, remote interleaved and remote blocked. In experiment 1 participants practised calculating volumes, split into three groups: blocked, interleaved, and remote interleaved. The blocked group practised calculating four types of volume questions based on the four different shapes. The standard interleaved group practised questions with the shapes interleaved but the task blocked (e.g. all finding the volume). The remote-interleaved group practised questions with volume questions interleaved with other topics such as fractions and exponents, this is similar to the intervention introduced in this study. The order and grouping of questions are demonstrated in Table 1.

Table *1*: The order of practice received by blocked, interleaved, remote interleaved and remote blocked groups in the Foster et al. (2019) study (p. 1091)

It was in Experiment 2 that the fourth group was introduced, 'remote blocked', whose practice was based on different topics but blocked, as demonstrated in Table 1 above. This group served as a comparison to the remote interleaved. Results for both interleaved and remote interleaved were positive, providing evidence for the use of interleaved practice over blocked practice.

Further evidence for the support of interleaved practice can also be found in those classroom-based studies, providing ecologically sound evidence to the benefits. In many of these studies, the content of the interleaving practice varies. Ziegler & Stern (2014) found that contrasting superficially similar, but conceptually different, material resulted in improved long-term learning. With a similar focus on contrasting the presentation of material, investigating whether task or topic should be interleaved, Rau Aleven & Rummel (2013), compare the questions included in the interleaved practice. In their

investigation, involving 158 participants of 10-12-year olds, the representation is interleaved (topics interleaved) or the task type is interleaved (using a diagram, filling in the blank etc). Rau, Aleven and Rummel (2013) found that “interleaving task types leads to more robust representational knowledge." (p. 109). Whilst this was a classroom experiment, thus making it more relevant to my study, it was carried out using educational technology (Cognitive tutors) rather than teacher led examples.

However, the same positive results have been found using a large, diverse sample (54 classes over four months) in a naturalistic setting, with teachers rather than educational technology (Rohrer et al., 2020b). In their randomised controlled trial, to assess interleaved practice, the participant groups either completed blocked assignments or interleaved assignments. The 'interleaved group' received blocked practice at the introduction of a topic (taught topic - practice a few) before going on to interleaved practice. Similarly, to ensure the test delay was consistent for both groups, they both received interleaved review before the test, resulting in the 'blocked group' receiving some level of interleaved practice. Therefore, as each group received some level of both blocked and interleaved, Rohrer et al. (2020b) acknowledge that their results demonstrate the difference between low and high interleaving.

Whilst these studies were in a more applicable setting, not all classroom-based studies showed statistical significance. Ostrow, Heffernan, Heffernan & Peterson (2015) conducted a classroom-based study with a sample of four teachers and 146 students (of similar age to those in this study). Although test scores showed positive effects for those who received the interleaved practice style over those who received blocked practice, this was not statistically significant.

Having considered the studies investigating interleaved practice, in both laboratory and naturalistic setting, it is important to address whether these benefits are because of interleaved practice. It should be considered that there are other learning strategies that are integral elements of interleaving.

2.2 Other learning strategies involved in interleaving

Given the importance placed on presenting strategies in a non-consecutive fashion for interleaving to be achieved, then it must be acknowledged that for that to happen those strategies must be spaced. Therefore, if interleaved practice is employed then spaced practice is also employed. Spacing the presentation of problems is a well-known and researched learning strategy producing powerful results (e.g. Yazdani & Zebrowski, 2006; Pyc & Rawson, 2007; Rohrer et al., 2014). Additionally, if learners are to access the strategy required for questions then there must be an element of retrieval. Bringing yet another powerful learning strategy to the forefront as an integral element of interleaving.

In fact, such is the power of spacing and retrieval practice that Dunlosky et al. (2013), in their commissioned evaluation of 10 learning strategies, found that spaced practice and retrieval practice (called Practice Testing in their evaluation) were the only two to receive a high utility rating. Whilst the focus of this study is on the use of interleaved practice, it is important to acknowledge and explore those covariate learning strategies. Indeed, as mentioned by Rohrer et al. (2020b) both learning strategies are recognised as covariates of interleaving:

> "the interleaving of practice problems in a course or text inherently incorporates the learning strategies of spacing and retrieval practice, each of which is an effective and robust learning strategy." (p. 41)

Thus, definitions for both spaced practice and retrieval should be addressed. Followed by an exploration into any studies which have sought to eliminate those other variables.

2.2.1 Spacing

Interleaved practice ensures spaced practice; when practice problems are interleaved the scheduling for each type of problem leads to inherently spaced practice. It is not possible to have one without the other and, therefore, if interleaved practice is adopted then it will follow that spaced practice is also taking place. Whilst a lot of the research pertains to conclude that interleaved practice is effective, there is recognition from authors of this other educational theory, spaced practice, intertwined within these studies (Taylor & Rohrer, 2010; Rohrer et al., 2014; Rohrer et al., 2020b). In fact, it has been

recognised as an “inherent benefit of interleaved mathematics practice" (Rohrer et al., 2015, p. 906). Earlier studies have perhaps failed to distinguish between spaced effect and interleaving effect, resulting in these confounding variables. Although Taylor and Rohrer (2010) introduce the concept of spacing as a fundamental part of interleaving, claiming that interleaving ensures that practice time for a skill is distributed. A point which is also supported by Rohrer et al. (2014) who argue that interleaved practice has two critical features: "problems of different kinds are interleaved (which requires students to choose a strategy), and problems of the same kind are spaced (which usually improves retention)." (p. 1324). Given this covariate nature, it is important to address the definitions of spaced practice.

When exploring the literature, the studies investigating spacing tend to use multiple terms for the concept of spacing the presentation of material. ‘Spaced’ and ‘Distributed’ are used interchangeably, as found here by Rohrer and Taylor (2006): "When practice is distributed or spaced, a given amount of practice is divided across multiple sessions." (p. 1209).

Similarly, Yazdani and Zebrowski (2006) define ‘Spaced Reinforcement’ as the presentation of newly taught material over an extended period, which they specify as “usually five or more non-consecutive class sessions” (p. 37). They compare this to ‘Massed Reinforcement’, whereby newly taught skills are applied many times, shortly after being presented to the learner, in a concentrated period. Here the separating factor is the need for reinforcement of new topics to be spaced over multiple timing sessions to secure the knowledge, in contrast to practice concentrated into one session until mastery achieved. By this account, mastery may be achieved in the short term but will not be secured in long term learning.

This definition of distributed vs. massed practice is supported by that of Pyc and Rawson (2007):

> "distributed practice involves multiple presentations of an item with intervening time and material between each, as opposed to massed practice in which all presentations occur immediately following one another." (Pyc & Rawson, 2007, p. 1917)

The benefit that spacing (distributing) topics has on memory is known as the spaced effect: “a memory advantage that occurs when study sessions are spaced apart in time” (Kapler, Weston & Wiseheart,

2015, p. 38). The spaced effect has been widely researched and has been described as a well-established finding in experimental psychology, yielding better memory than massed practice over a range of methodology (Pyc & Rawson, 2007). Clearly spaced practice is a powerful tool for material retention.

Indeed, there is substantial evidence to support repetition through spaced practice rather than repetition through blocked practice (Bjork, 1979; 1994; Cepeda et al., 2006; 2008; Pashler, Cepeda, Carpenter & Rohrer, 2007). However, it should be acknowledged that much of this research is conducted in ways that lack validity in schools, often using adult participants, conducted in laboratory settings, and focusing on simple verbal tasks.

Therefore, in narrowing down the focus on spaced practice within educational research, studies have found that distributed practice has a better effect on memory than massed practice (Cull, 2000; Roediger & Karpicke, 2006; Pyc & Rawson, 2007). Pyc and Rawson (2007) developed previous research into studying the number of trials assigned to each item, completing a study based across two experiments, with 227 undergraduate participants learning Swahili-English vocabulary words pairs. The study, using robust methods, found evidence in support of a drop-out schedule of spaced practice:

> "Overall, the present research provides initial evidence that a dropout schedule can be more efficient than the conventional schedules that have been widely studied in previous research." (Pyc & Rawson, 2007, p. 1924)

Whereby the dropout schedule involved practice of a given item being discontinued once the learner had reached the performance criteria. Implementing distributed practice on a schedule alongside restudy yielded a higher performance than restudy or retrieval alone. Findings were also confirmed in Cull's (2000) study on word recall. Although both studies were produced in a laboratory setting and were not tested on complex material, or over a long period of time.

Continuing to narrow the focus further, specifically, in mathematics for relevance to my study, the body of research is smaller. Studies are fewer, but nonetheless still widely reported spanning over decades. In multiplication facts (Rea & Modigliani, 1985) in statistics (Smith & Rothkopf, 1984), in

geometry (Rohrer & Taylor, 2006), mathematics fact fluency (Schutte et al., 2015), and even more recently mathematical vocabulary (Petersen-Brown et al., 2019). In a combination of both interleaving and spacing, Rohrer et al. (2014) designed a study that would incorporate interleaved practice but ensuring that practice of skills was distributed in a delayed manner across the different assignments. They recognise the limitations of the large effect observed in their study (d=1.05), acknowledging the role of spacing in the contribution to the effect size.

Furthermore, distribution of practice has also been found to have positive effect on mathematics homework. Yazdani and Zebrowski (2006) conducted a study, with strong ecological validity due to its classroom setting - following the curriculum, involving two experiment groups and one control group (pre-test determined they were equivalent initially). The six-week experiment focused on spaced reinforcement in homework, designed on a dropout scale e.g. to reinforce Topic A the experiment group would receive a homework on Day 1 of which 30% was Topic A, Day 2 20%, Day 4 15% and so on. On day 7 the homework would contain 15% topic A, down to day 20 homework containing 5%. The sliding scale for practice of this topic presents similarities to both Rohrer et al.'s (2014) sliding scale and the, earlier described, Pyc and Rawson (2007) drop out schedule, in that it reduces the amount of that topic presented rather than where it falls in the practice. They found, using an ANCOVA analysis and Tukey-Kramer test, that spaced practice for homework was "significantly superior in engendering students' meaningful long-term memory." (Yazdani & Zebrowski, 2006, p. 38). The notion of engendering pupils' long-term memory is of key importance. The homework element of this study was of interest as this distribution is akin to the current homework set by the department as stated in the introduction.

In short, in real-world scenarios interleaved practice is going to involve the element of spaced practice. Lessons are usually spread out throughout the week and therefore any classroom-based studies are inherently going to incorporate a spaced practice element. Alongside this, if the presentation of problems is to be non-consecutive, as deemed definitive of interleaved practice (Foster et al., 2019, Rohrer et al., 2020b), then they must also be spaced. Given the evidence supporting the superiority of

spaced practice, as discussed above, this can only provide further support for the use of interleaved practice.

2.2.2 Retrieval

This second and significant component of interleaved practice, and specifically the type of interleaved practice used in my study, is the concept of retrieval practice. Learners must retrieve and assess information before embarking upon an appropriate strategy to answer a question (Rohrer et al., 2020b). It is not enough to simply reuse the technique employed in the previous question the learner had faced (as takes place in blocked practice).

The retrieval process will influence learning indirectly by strengthening subsequent encoding, and directly; the act of retrieval itself enhances learning. Recently, the focus of educational research has been on the links between active retrieval and the promotion of meaningful learning (Dunlosky et al., 2013).

Karpicke & Grimaldi (2012) define the process of retrieval:

> "Retrieval processes are involved in all situations in which knowledge is expressed, including situations where learners must produce the answer to a factual question, explain a concept, make an inference, apply knowledge to a new problem, and produce creative and innovative ideas." (p. 401)

For mathematics, learners must discriminate between problems; they need to employ a strategy and in doing so, engage in retrieval processes that bring knowledge to an applicable level. Learning has been proven to be strengthened by this challenge (Adescope et al., 2017).

As with spaced practice, a plethora of research supports the use of retrieval practice. Whilst, this evidence comes primarily from laboratory studies (Adescope et al., 2017), positive effect sizes have been found to be equivalent in classroom studies (e.g. Agarwal et al., 2014; Agarwal & Bain, 2019; Rawson & Dunlosky, 2012; Dunlosky et al., 2013). Specifically, recent research in classroom settings proves this impact on learning for diverse subjects and educational settings, as well as time delays. Retrieval, as a learning strategy, is an area of research that is progressing rapidly, extending in areas such as mathematics and science (Dunlosky et el., 2013).

Indeed, in an ecologically valid experiment, McDaniel et al. (2013) examined fact and application questions used in retrieval practice in their second experiment, in a science classroom. They found that, regardless of format, final test performance was significantly increased with the use of retrieval practice; learners engaging in complex retrieval practice found a greater benefit than starting with basic facts and definitions. The effect size found by McDaniel et al. (2013) is large for a classroom experiment but could be pertained from a combination of strategies: retrieval, testing and spacing.

As evidenced above, the act of retrieval is a powerful tool. Indeed, this benefit of retrieval practice is not limited to test results and has been found to reduce examination anxiety as well as improve learning (Agarwal et al., 2014). In considering this, an exploration into the optimal use of retrieval should be made. As revealed by Rawson and Dunlosky (2012), the most efficient and effective retrieval process comes from a correct retrieval during initial presentation, followed by three to four subsequent presentations. In addition to this, is the spacing of the retrieval, whereby the retrieval becomes increasingly difficult as the time since last exposure increases (Pyc & Rawson, 2009). However, recent evidence indicates that to get maximal results, time between exposure should be expanded over time (Kang, Lindsey, Mozer, & Pashler, 2014). This principle of a longer period between exposure, is like the dropout schedule proposed for spaced practice (Cull, 2000; Pyc & Rawson, 2007).

All things considered, it is clear that the coherent nature of spacing, retrieval and interleaving act in parallel alongside each other. It is easy to question whether any effects found for interleaving are, indeed, due to interleaving or because of spacing and retrieval.

2.2.2 Isolating Interleaving

Acknowledged above, two key learning strategies that are inherently ingrained in interleaving are spacing (predominantly) and retrieval. Given that those earlier definitions of spacing and interleaving appear on the surface very similar in their design, it feels important to seek whether a separation of the two is made in the literature and whether any studies have sought to counter those variables.

The level of questioning has been identified as one of those separating factors, between spacing and interleaving, found at both the practice phase and testing phase. During the testing phase, novel items are presented with interleaving rather than memory of the original item as in spacing (Firth, Rivers & Boyle, 2019). In practice stages, spacing will involve restudying the same material, whereas interleaving may involve the presentation of different examples.

On that basis, it can be said that if interleaving happens in a naturalistic setting then spacing must occur, as the mixing of strategies inherently leads to the spacing of strategies over time. However, if spacing occurs then this does not inherently mean interleaving has happened in the learning of a new concept. A topic may have been taught and practiced in a blocked manner but then reinforced using spacing.

In appreciation of these factors, various studies sought to equate the effects of interleaving with the degree of spacing (Taylor & Rohrer, 2010; Kang & Pashler, 2012). This came from the recognition that many studies, accrediting the impact and benefits to interleaving, could have, indeed, reflected the spacing rather than interleaving. Kang and Pashler (2012) replicated Kornell and Bjork's (2008) study of the presentation of artists, factoring out the spacing effect by introducing filler items to ensure that the spacing of each group was constant. They found that interleaved conditions still produced a positive effect despite removing the spacing element. Thus, providing evidence for the benefits of interleaving alone.

Additionally, these benefits have also been replicated in mathematics education (Taylor & Rohrer, 2010). Firstly, Taylor & Rohrer (2010) equated the spacing element in their study. Participants, of a similar age to my study, received the interleaved practice all within one session or the blocked practice all in one session. Pupils then returned the next day for a test where it was found that, even though the degree of spacing was held constant, the test scores more than doubled for the interleaved group. Although, they claim that in a natural classroom setting it is not feasible to have fully interleaved and blocked practice groups, as there must be an element of taught topic at the beginning of the lesson.

To summarise, interleaved practice is where problems are mixed in a specific way. The type of question, topic, skill, and strategy required, presentation and testing are done so in a non-concurrent nature, so that the problem cannot be solved by the same strategy as the previous question. As an inherent element of interleaving, distributed practice is where the focus is on the spacing of practice; practice is done so in short bursts over a long period of time as opposed to all in one lesson. In each of these, the act of retrieval of past knowledge for future application will be required. The reason for addressing these elements is because of the coherent nature of them and to recognise that the effects of interleaving, as discussed below, may not be as a result of interleaving in isolation; interleaved practice will naturally involve an element of spacing and retrieval.

2.3 The effects of interleaving

As has been discussed, the benefits of using interleaved practice in both non-educational and educational settings have been widely observed. Therefore, it seems necessary to explore further what effects have been found when using interleaved practice. The common effects found are improved ability to discriminate between problems and associate the relevant strategy, and improved test results (e.g. Mayfield & Chase, 2002; Taylor & Rohrer, 2010; Rohrer et al., 2014; Foster et al., 2019). All of which are to be discussed in this section. The changes in these effects over time will also be addressed.

2.3.1 The ability to discriminate between problems

When problem solving, mathematicians must learn to be able to discriminate between problems, this is relevant to every level of mathematics but especially school level learners who are novice mathematicians. In addition to this, pupils need to be able to discriminate between both superficially similar and dissimilar problems and then associate the problem with an appropriate strategy, whereby “students often find the choice of strategy to be more challenging than the execution." (Rohrer et al.,

2015, p. 900). This difficulty can be worsened by superficially similar problems requiring different strategies to solve them.

It is argued that most assignments are arranged in a way that simplifies the solution to a problem, or guides the learner towards a method before even reading the problem, thus reducing the pupil's need to identify a strategy themselves (Rohrer et al., 2014; 2015; Foster et al., 2019). The discrimination between problems is removed when students are directed to the correct strategy. Blocked practice presents this issue, being taught a lesson on trigonometry, for example, and then being faced with several problems requiring the use of trigonometry, removes the need for identification of employing trigonometry. Pupils are instructed of the strategy before they have even begun the task. The problem with this being that "this crutch is usually not available to students when they are tested." (Rohrer et al., 2015, p. 900). This provision of the required solution is weakening students' ability to problem solve when asked to do so on their own.

As this difficulty can be experienced by mathematicians at any level, it is important to try to establish ways in which this strategy selection can be improved and refined. Interleaved practice has been found to do just that (e.g. Rohrer et al., 2015; Foster et al., 2019). Foster et al. (2019) attribute the effects of interleaving to the 'discriminative-contrast hypothesis', which they claim improves problem solving "because instances of that problem type are practiced in close proximity to instances of different problem types from a similar domain." (p. 1088). The benefit of being able to discriminate between superficially similar problems has resulted in improved performance in tests (Kang & Pashler, 2012; Kornell & Bjork, 2008). Rohrer and Taylor's (2007; 2010) experiments both demonstrated evidence to support the 'discriminative-contrast hypothesis'.

Whilst the benefits of interleaved practice can be accredited for problem strategy association and, therefore, improved problem solving, it poses the question of whether this benefit can be experienced for superficially dissimilar problems. It has been argued that, if the sole benefit of interleaved learning is the ability to discriminate between problem types, then this benefit should

not be experienced for dissimilar problems (Rohrer & Taylor, 2007; Taylor & Rohrer, 2010). On that basis, Foster et al. (2019) question the use of dissimilar problems within interleaved practice:

> "...when the problem types are highly dissimilar then participants would not need to learn what aspects of each problem are relevant to assigning it to a problem type." (p. 1090)

Despite this, the 'remote interleaved' group (using dissimilar problems in interleaved practice) in Foster et al.'s (2019) study, see Table 1, outperformed the blocked group and standard-interleaved group (using similar problems in interleaved practice). Although these results were not significant compared to standard-interleaved (d=0.36), the results were significantly greater compared to the blocked group (d=0.56) (Foster et al., 2019). Consequently, providing evidence to contradict their initial doubt of the use of dissimilar problems.

The evidence for the use of dissimilar problems in interleaved practice is not limited to Foster et al.'s (2019) study. In two classroom-based studies, Rohrer, Dedrick & Burgess (2014) and Rohrer, Dedrick and Stershic (2015) both found positive effects of interleaved practice. Different to other studies (e.g. Rohrer & Taylor, 2007; Taylor & Rohrer, 2010) which used highly similar problems in their interleaved practice, both studies ensured that the types of practice questions included were superficially dissimilar. There was a statistically significant difference for those pupils who received the interleaved practice, credited to the improved ability to choose the correct strategy:

> "As argued here, these benefits arise because interleaved practice provides students with an opportunity to learn how to choose an appropriate strategy (or learn that they cannot do it)." (Rohrer et al., 2015, p. 906)

The positive effects found in these studies, with the use of superficially dissimilar problems, lend themselves to contradict the 'discriminative-contrast hypothesis' (Foster et al., 2019).

To summarise, improved discrimination between problems, as well as strengthening of the association between different problems and their corresponding strategy, can be made using interleaved practice. The use of dissimilar problems in interleaved practice was initially disputed by Foster et al. (2019), who believed that this would reduce the problem strategy association. However, results from their

experiment (also see Roger et al., 2015) supported the use of both similar and dissimilar problems in the practice. This suggests the positive effects of interleaved practice can be found using both superficially similar and dissimilar problems.

2.3.2 Improved test performance

As demonstrated above, interleaved practice has been shown to improve learners' ability to discriminate between problems, improving their ability to assign strategies to problems. This ability to select the appropriate strategy has been shown to improve performance in tests (Birnbaum, Kornell, Bjork, & Bjork, 2013; Carvalho & Goldstone, 2015; 2019; Kang & Pashler, 2012; Kornell & Bjork, 2008). One study even found that test results of the interleaved group was triple that of the blocked practice group (Rohrer & Taylor, 2007).

In multiple laboratory studies, interleaved practice has been found to produce significant benefits to test results (Mayfield & Chase, 2002; Taylor & Rohrer, 2010; Foster et al., 2019). Mayfield and Chase (2002) found that interleaved practice produced better test scores in the learning of algebraic rules. Interestingly in this study, the interleaved group received their skill practice in an interleaved cumulative fashion; learning two skills with mixed practice, before adding in a third skill with three skill mixed practice, continued up to five skills. This methodology is one which has been advised will produce maximal benefits of interleaved practice (Dunlosky et al., 2013). The practice of skills was followed by a test whereby all five skills were assessed. Again, the interleaved group outperformed the massed group (97% accuracy compared to 85% accuracy). This concludes that the students' ability to solve novel algebra problems was substantially boosted by interleaving. However, Mayfield and Chase (2002) recognise the limitations of their own study, acknowledging that control regarding timing, and exposure to variables, was limited due to logistical constraints. As discussed earlier, the covariate nature of spacing and interleaving should be recognised.

Producing similar superior test results, both Rohrer and Taylor (2007) and later Le Blanc and Simon (2008) used interleaving for the learning of volumes of solids. In addition to investigating test performance, Le Blanc and Simon (2008) also measured participants' ability to estimate their performance. Participants receiving interleaved practice, alongside getting superior test results, were able to better estimate their performance (both before the test and after). Similarly, in Taylor and Rohrer's (2010) study test performance was significantly higher for the students who faced interleaved practice over those that faced blocked practice (77% accuracy compared to 38%).

The positive effects found in the aforementioned studies have held over time and even more significantly, in classroom settings. In Rohrer et al.'s (2020b) randomised controlled trial, discussed earlier, groups either completed blocked assignments or interleaved assignments and at the end both groups completed a mixed review assignment, similarly positive test results were produced, whereby the interleaved groups outperformed the massed groups with a large effect size (61% to 38%, d=0.83).

2.3.3 Effects over time

It has already been discussed that spaced study sessions result in better learning than massed study sessions and this is an integral aspect of interleaving (e.g. Cull, 2000; Pyc & Rawson, 2007; Schutte et al. 2015). However, the timing of testing in both spaced and interleaved practice has been measured over a great range. Whilst some have found benefits immediately and at a delayed period (e.g. Rohrer et al., 2015), others found no immediate benefits but delayed testing benefits (e.g. Nazari & Ebersbach, 2019), and some found benefits both immediately and increasing over a delay (Rohrer & Taylor, 2006; Rohrer et al., 2015).

Suffice it to say, findings vary, which leads to the suggestion that teachers may need to wait to see results of implementing interleaving. Not all studies have found immediate results; Rohrer et al. (2015) found that the effect of interleaved was smaller on the shorter test delay. Similarly, Ostrow et al. (2015) failed to find a positive effect of interleaving using a test delay of 2-5 days. However, in contrast to this Taylor & Rohrer (2010) found positive effects from interleaving after test delays of just one day.

Despite benefits not always being found in the immediate, there has been evidence to suggest benefits after a delay. Rohrer and Taylor's (2006) study produced findings that small intervals in test time resulted in limited benefits, however, a large spacing effect was observed for the 4-week test. The same result was found in a more recent study by Nazari and Ebersbach (2019). Using a Bayesian analysis of the performance, no effect was found for the test taken two weeks post practice, however, strong evidence was found for the delayed test six-weeks later. Indeed, exploratory analyses indicated that those whom benefitted most from the distributed practice were the mid-range performers, providing further support for the inclusion of distributed practice as an integral part of interleaved in my intervention (the study sample are a middle set).

Furthermore, in a combination of immediate and delayed effects, there is an argument that the benefit of interleaved practice can not only be found immediately but has also been proven to grow over time. In their study, Rohrer, Dedrick & Stershic (2015) tested students at different intervals, one day and thirty days after the review. They assessed the effects of interleaved practice in experiment one by comparing it to blocked practice, then they assessed the effects of overlearning in experiment two by varying the amount of practice of each problem. As the two elements of the study were orthogonal, it was purposed to find whether one, both or neither of these strategies was beneficial to pupils learning. They claimed that "the study is also the first to demonstrate that the test benefit of interleaving does not diminish over time and perhaps grows larger." (Rohrer et al., 2015, p. 906). They announced positive outcomes from interleaving, but they also acknowledged consideration of the effect spacing out the tests could have had.

Finally, in consideration of the short intervention times presented in some of the previously mentioned studies it needs to be considered whether the same results can be found for longer studies. In two studies with those longer time intervals, by Rohrer et al. (2015) and Rohrer et al. (2020b), students received mostly interleaved or mostly blocked practice over several months, before completing a review assignment. One month after the intervention, learners faced an unannounced

test and those who had received the higher dose of interleaving produced much higher test scores (Rohrer et al. 2015; Rohrer et al., 2020b).

Clearly there is substantial evidence supporting the use of interleaved practice, determined by positive outcome measures; improved test results with the effect strengthening over time, and learners being better able to discriminate between and associate relevant strategies to problems. Then, for long term learning and retention in mathematics, interleaved practice should be considered. Despite this, the strength of a learning strategy should not just be measured on test performance but the experience for the participants involved should also be discovered to help to inform practices.

2.4 The experience for participants

Having explored the many benefits of interleaved practice, it is important to explore what the experience of implementing it is. This section will explore various aspects of the experience for both the teacher and the learner.

2.4.1 For the teacher

Whilst it is important to improve attainment, as interleaved practice has been demonstrated to do, it is also important to understand whether it is feasible for teachers to implement the ideas.

Dunlosky et al. (2013) claim in their monograph, in which they offer recommendations about the relative utility of learning strategies, that introducing interleaved practice is feasible for teachers:

> "teachers and students could integrate interleaved practice into their schedules without too much modification." (p. 44)

Despite this claim, very few of the studies assess the feasibility of interleaved practice for the teachers. In their randomised controlled trial of mathematics learning with an interleaved intervention, Rohrer et al. (2020b) sought to establish this feasibility in their study, whilst also assessing effectiveness under more suitable conditions. Their study, based on a large diverse sample (54 classes over four months) in a naturalistic setting, has two aims: 1. Assess interleaved practice under naturalistic conditions, 2.

Evaluate the feasibility for teachers implementing the practice in their classrooms. Before learning the results of the study, teachers expressed a favour towards interleaved practice: “a positive effect was found for each of the 15 teachers" (Rohrer et al., 2020b, p. 48). Teachers agreed that interleaved practice was good for improving test scores and was appropriate for low achieving as well as high achieving learners. With most teachers reporting that they liked interleaved practice and would be recommending it to other teachers, this suggests that not only is interleaved practice effective, but it is also feasible.

One such factor which will affect feasibility for teachers, is the resources available to them. Many teachers will recognise the standard method of practice within a lesson; the practice problems are based on the most recent (often just taught) strategy, blocking the type of problem. This presentation of work in a blocked fashion, practised with blocked resources, certainly presents itself in the department I am a member of. As discussed earlier, this can result in a lack of ability to learn to associate correct strategies with their problem. It is, therefore, important to consider what role resources play in the implementation of interleaved practice.

Predominantly, taught lessons, resource websites and textbooks are dominated by blocked practice. Thus, these resources, by using similar problems and delineated topics, potentially promote the practice of overlearning, with no long term benefits (Rohrer & Taylor, 2006; Rohrer et al., 2015), and deficient processing (Rohrer et al., 2014; 2015; Foster et al., 2019). Leading to the current format of mathematics textbooks, or the linear approach teachers may use when using those, to be questioned.

Consequently, many authors have called for the need for textbooks to contain spaced and interleaved practice (Rohrer & Taylor, 2006; 2007; Casey et al., 2018; Rohrer et al., 2020a). Rohrer and Taylor (2007) praise textbooks which provide practice problems in a shuffled format, whereby questions are systematically distributed throughout a textbook in decreasing frequency. They claim that this "intrinsically ensures that the problems within each practice set include a mixture of different

types"(p. 482). Their suggestion, following their findings, was that learners should be presented with only three to four questions in succession of any problem type (Rohrer & Taylor, 2006).

Given the lack of interleaved resources readily available to teachers, the introduction of more interleaved resources would mean that interleaving in lessons would be even more feasible for teachers. In addition to that, it would provide learners with the ability to experience an interleaved presentation of problems in their practice outside of the classroom.

2.4.2 For the learner

Desirable difficulties

Many benefits are believed to be taken from a teacher who creates desirable difficulties for their pupils, whereby these desirable difficulties are aimed to encourage higher order thinking in mathematics but also to enhance transfer of knowledge to other problems and long-term retention (Bjork, 1994; 2013). There are various systems that will intensify the learning process for the pupil, hence the desirable difficulty notation, and these are thought to be the driver in that desirable, long-term memory retention. The powerful effects of these desirable difficulties have formed the focus of many cognitive psychology research projects (e.g. Carpenter et al., 2008; 2012; Dunlosky, et al., 2013; Rohrer et al., 2015; Toppino & Gerbier, 2014).

These desirable difficulties have certainly been experienced by the participants of many interleaved studies (Rohrer & Taylor, 2007; 2009; Rohrer et al., 2020b). Whilst the outcome of interleaved practice interventions has been found to have positive effects in final tests on learning formulae, problem solving, and test performance, the learners have experienced difficulties during the practice testing phase (Rohrer & Taylor, 2007; 2009; Rohrer et al., 2020b). Rohrer and Taylor's (2007; 2010) studies are an example of this, despite test results ending with the interleaving group significantly outperforming the blocked group, their accuracy was impaired during practice. In their 2010 study, as can been seen in Figure 1, the blocked practice group outperformed the interleaved practice group in all stages of practice, with both partial and full problems.

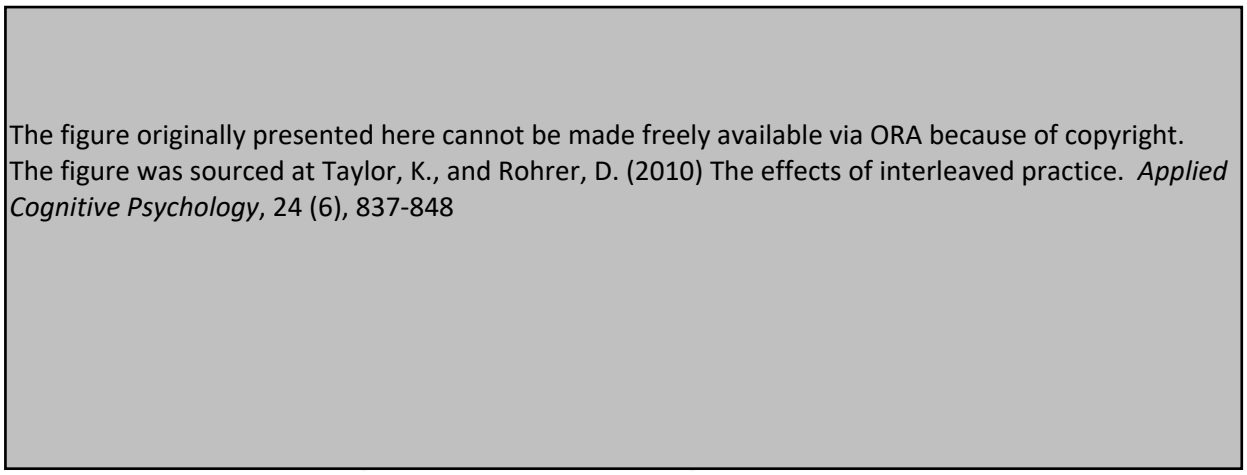


Figure 1: The performance of the blocked group compared to the interleaved group throughout the practice stage (to include both partial and full problems) and at the post-implementation test (reprinted from Taylor and Rohrer, 2010, p. 843).

Despite Figure 1 demonstrating the blocked group outperforming the interleaved group, the blocked group then significantly underperformed in the test when compared with the interleaved group (38% to 77%); as such, providing evidence that the interleaved practice provides the opportunity for desirable difficulty. Those difficulties experienced during the practice phase clearly enhance the transfer of knowledge during the test phase. Although the impact on long term retention is not measured here.

Providing qualitative support to the argument, this desirable difficulty was also witnessed by teachers in Rohrer et al.'s (2020b) study. Several weeks after the test, the teachers were given a survey to complete (before they were then informed of the results of the study). The survey found that teachers thought that their students found the interleaved practice harder and more time consuming. Whilst this was not voiced by the students themselves, it was recognised by the teachers.

Not unlike desirable difficulty, the element of productive failure has been discussed (Kapur, 2008; 2014), suggesting that if learners face early difficulties in problem solving, this can be beneficial for long-term learning. Furthermore, difficult retrieval practice may benefit test performance to a greater extent than easier retrieval practice (Bjork, 1994; Pyc & Rawson, 2009). All of which could be contributors to the difficulties and challenges faced by learners during the practice phase, resulting in positive outcomes in the end.

Engagement

The use of interleaved practice may engage learners' attention more effectively than completing blocked practice. Different practice structures experienced in lessons by the learner could result in differing engagement levels. Lack of engagement in lessons may be recognised as mind-wandering, as studied by Metcalfe & Xu (2016). In their category-based learning study (artist depiction), they found that mind wandering was significantly higher for those during blocked conditions compared to that experienced during interleaved conditions. This was especially the case as the task progressed; there was little difference in self-reported mind wandering at the beginning of the study compared to a larger amount towards the end.

Perhaps another indicator of reduced attention in practice is reduced study time, as experienced by Wahlheim et al. (2011). The presentation of perceived trivial tasks in both studies may have resulted in reduced engagement over time. This provides further evidence for the use of interleaved practice in lessons, as "interleaving alters the pedagogical demand of a mathematics problem" (Rohrer et al., 2014, p. 1328). This pedagogical demand may result in the learner's engagement increasing, thus reducing mind wandering. If the difficulty learners experience (as discussed earlier in desirable difficulty) during interleaved practice better prepares them for future application, then learners might engage on a more attentive level with feedback presented during that practice time. Thus, further improving engagement with all aspects of learning in the lesson.

Although lesson engagement and motivation are big factors of pupil participation, so too are examination anxiety and the pressure learners feel once faced with novel problems and the demand to retrieve the information. As explained in the introduction, this was certainly a common occurrence in my setting. Retrieval practice, an earlier defined integral element of interleaving, has been proven to support the reduction in test anxiety (Agarwal et al., 2014; Agarwal, 2019). In their classroom-based study on retrieval practice, which was one of the first in retrieval practice to assess the experience for the participants, Agarwal et al. (2014) found that 72% of students reported that the use of retrieval resulted in them feeling less nervous about assessments. They recommend retrieval practice be used

by teachers to both reduce test anxiety and improve learning (Agarwal et al., 2014). Given that interleaving provides the opportunity for retrieval, this would suggest that, learners challenged with retrieval of strategies should benefit from reduced examination anxiety.

2.5 Summary of Literature

The evidence detailed in the literature review makes a clear case in support of the use of interleaved practice within the mathematics classroom. Whilst there is still not a considerable amount of research in the field, the presentation of both laboratory and classroom-based research suggests that interleaved practice is superior to blocked practice (e.g. Dunlosky et al., 2013; Rohrer et al., 2014; 2015; Foster et al. 2019).

Interleaved practice may provide the pupils with the opportunity to learn how to discriminate between problems, thus allowing them to learn how to identify what strategy is required, enabling this strategy and problem association to happen in a test (e.g. Rohrer et al., 2015). Additionally, the benefits found may also arise from the two integral aspects of interleaving; interleaved practice provides an opportunity for spaced practice and for retrieval. When study and restudy of material is separated with a delay, learning is benefitted (e.g. Schutte et al., 2015; Peterson-Brown et al., 2019). The act of retrieval, especially when spacing is expanded over time, is also beneficial for learning (Adescope et al., 2017). Despite this, there have been studies that have equated the spacing and still found positive effects from interleaving alone (Taylor & Rohrer, 2010; Kang & Pashler, 2012), although this is not thought to be possible in naturalistic settings (Taylor & Rohrer, 2010).

The outcome of implementing interleaved practice and the effect this has on test results has been broadly researched, proving to produce superior results to blocked practice (e.g. Rohrer et al., 2014; 2015; Rohrer et al., 2020b). However, a gap in the research comes in the reported experience of the participants.

Therefore, shaped from what I have learnt in the literature, these findings suggest that current standard practices in the department teaching are not optimal. Thus, providing the foundation and

motivation to evaluate this learning intervention within my own school setting, exploring whether superior test results are found, both at an immediate test and a delayed test. Concurrently, given the gap in the research, it is of importance to me to establish the experience from the learners' perspective.

As such, the research questions formed out of this literature review are as follows:

1. What effects did the use of interleaved practice in lessons have on test results?
2. What were pupil perceptions of interleaved practice in lessons?
3. How do the effects and perceptions change over time?

The methodology for which, will now be discussed.

3. Methodology

3.1 Overview

The intervention was carried out with my Third Form (Year 9 – age 13-14), middle attaining, teaching class, in their mathematics lessons. The intervention was designed to explore the use of interleaved practice within lessons, where I was the teacher implementing the intervention.

Pupils were taught a skill, followed by four questions practising that skill, with the remainder of the lesson spent completing interleaved practice (of previously taught topics). The intervention was over a three-week period. The parallel set to my class (same year group, middle attainers), taught by another teacher, continued to be taught as is normal practice in the department, using blocked practice (teach a skill – practice a skill). At the end of the three week intervention, as described above, the two classes sat the same test (written by the Head of Department to remove bias and discussed with myself and the other teacher). The results of this test were analysed, and comparisons made.

Qualitative data was collected at three points post-intervention. Firstly, a survey of the intervention class was carried out immediately after the intervention to gather pupil perceptions of interleaved

practice. Secondly, a group interview was carried out with three pupils from the intervention class. Finally, a survey of the intervention class was carried out four weeks after the intervention to try to establish how pupils perceived interleaved practice could be used in remote lessons.

As part of the normal procedure at the School, for end of year examinations for this year group, pupils sit a PUMA (Progress in Understanding Mathematics Assessment) test. This PUMA test was due to take place four weeks after the intervention, where the plan was to analyse and compare the two classes PUMA tests. The purpose of this further collection and analysis of data, was to try to establish whether there was a difference between an immediate and a delayed test. However, this was not possible due to the sudden closure of schools in the UK because of the COVID-19 pandemic. Therefore, an analysis of this effect was not possible. The changes in methodology, because of the School closure, is described in the next section.

3.2 COVID-19 Impact

Change in methodology

During the study, schools faced closure due to the COVID-19 pandemic. This had an impact on the data that was collected and thus the study was altered part way through. An overview of the changes made to the study can be found in Figure 2.

Figure *2*: Pre/Post COVID-19 timetable. The timelines of the originally planned intervention (PRE-COVID-19) and subsequently changed timeline of the intervention (POST-COVID-19)

As demonstrated in Figure 2, the second post-intervention test had to be removed as this fell within school closure and it was decided, alongside the Head of Department, that it was not ethical to subject the pupils to further testing, therefore, the PUMA test was cancelled. Subsequently, the survey following the second test did not happen. Thus, removing the ability to include data of a delayed test and survey within this study. The delayed test and survey would have provided the opportunity to establish whether there was any greater improvement in test results from a delayed test, as found in the literature (Rohrer et al., 2015; Rohrer et al., 2020b).

Due to timing restrictions, the individual interviews also had to be changed to group interviews.

Change in research questions

Following the literature review I formed three research questions:

1. What effects did the use of interleaved practice in lessons have on test results?
2. What were pupil perceptions of interleaved practice in lessons?
3. How do the effects and perceptions change over time?

The closure of schools changed my ability to fully assess the change in impact on test results over time as I was not able to include another assessment.

After analysing the data collected from the post-intervention test and survey, seeing the pupils positive perceptions of interleaved practice in the initial phase, I adapted the third research question and remaining phase of the study to explore how best to use interleaved practice during closure. The three research questions addressed and analysed in the study, given the current circumstances were as follows:

1. What effects did the use of interleaved practice in lessons have on test results?
2. What were pupil perceptions of interleaved practice in lessons?
3. How can interleaved practice be used in remote learning?

Upon the physical closure of the School site, the School moved immediately to the delivery of full 50-minute lessons via video call, resulting in remote mathematics lessons. Therefore, this change in the final research question allowed me to establish pupil perceptions of the inclusion of interleaved practice in these remote lessons.

3.3 Collaboration

The main process for collaboration was with the teacher of the parallel set and the Head of Department. The teacher of the parallel set and I discussed and agreed on the topic being taught during the intervention phase, in line with the scheme of work this was to be simultaneous equations and the beginnings of probability. We collaborated to ensure that both classes were taught the same strategies and skills for those topics, using the same method. We also decided on and agreed together the topics to be included in the test. The topics to be included in the interleaved practice were discussed at a department meeting, whereby various members of the department gave ideas for potential problems and suggestions for different ways in which the skill can be tested. To ensure that the test that the two classes sat was as unbiased as possible, the Head of Department wrote the test

(based on the topics agreed) and then sent this to both teachers to be agreed. The test papers were also cross marked, collaboratively.

The coding of the quantitative data from the interviews and questionnaire was discussed, agreed, and therefore altered accordingly with the Deputy Headmaster. This discussion involved reading the transcript and asking whether the coding that I had assigned to the various comments, was agreed.

Once notice of school closure was received, the change in plan for the project was discussed with the Head of Department and Deputy Headmaster.

3.4 The Intervention

3.4.1 Participants

The intervention was carried out with my current Third Form (13-14-year olds) teaching set, containing 19 middle attainers. The set was chosen as they are one of my teaching classes in Key Stage Three with a comparable parallel set, taught by another teacher. In collaborating with the Head of Department, the sets were agreed as comparable based on their baseline data, taken from the previous term's assessment and their most recent PUMA (Progress in Understanding Mathematics Assessment) assessment.

Pupils in my teaching class (the intervention group) were introduced to interleaved practice. The parallel set continued to be taught in a blocked style as it would normally. Inclusion for the sample was as part of normal teaching lessons. The results of the test for both the intervention group and the parallel set were used for a comparison.

All qualitative data collected in this study was taken from the intervention group. The first questionnaire was sent to the 19 pupils in my class. The group interview was conducted with a small group from my class on a voluntary basis. The final questionnaire was sent online to my class.

3.4.2 Intervention design

Pupils in my class were taught a skill, practised four questions of that skill, and then received practice of previously taught topics interleaved, as detailed in the intervention interleaving plan in Table 2.

The new teaching topics for the duration of the intervention were simultaneous equations and probability. Pupils completed four current topic questions and then moved on to the interleaved practice. The choice for the blocked four questions was based on Rohrer and Taylor's (2006) suggestion that pupils should be presented with only three to four question of the same problem in succession, advice which has been used in previous interleaved studies for mathematics (Rohrer et al., 2014; Foster et al., 2019). As can be seen in the Lesson 1 column in Table 2, the first skill being taught was solving simultaneously when the coefficient of one of the variables was the same (using the elimination method). The practice session for the lesson was then as follows: four questions of simultaneous equations with the same coefficient for one of the variables, followed by 11 mixed topic questions to include topics such as: constructions, equation of a line and inequalities.

Once a new skill was learnt it was then integrated into the interleaved practice in the subsequent lessons. For example, the interleaved practice in lesson six contained three different simultaneous equations questions (each with a different strategy required) integrated into the other mixed topics.

An example of one of the interleaved practice sheets used in one of the lessons can be found in Appendix A.

Lesson	1	2	3	4	5	6	7	8	9	10
Main Teaching Topic	Simultaneous. Coefficient of x same	Simultaneous. Same coefficient, negative y	Simultaneous. Different coefficients.	Simultaneous. Different coefficients. Including	Solving simultaneous equations graphically	Solving simultaneous equations by	Basic prob	Drawing tree diagrams	Drawing tree diagrams (w/o replacement)	Using tree diagrams
Question										
1	14	14	15	15	16	17	18	19	19	20
2	14	14	15	15	16	17	18	19	19	20
3	14	14	15	15	16	17	18	19	19	20
4	14	14	15	15	16	17	18	19	19	20
5	6	3	8	5	2	10	14	17	15	5
6	9	2	7	3	9	4	7	3	9	18
7	2	8	10	6	15	8	5	15	16	10
8	4	5	6	14	7	14	9	10	8	11
9	8	7	1	4	8	6	17	11	14	15
10	10	11	9	11	1	5	6	2	18	1
11	7	1	14	2	10	16	11	5	4	17
12	5	9	2	10	6	11	3	16	17	9
13	13	6	4	7	3	7	15	1	5	6
14	11	4	11	1	14	2	16	7	3	14
15	3	10	3	9	11	15	8	6	1	16
16			5	8	4	1	4	14	10	2

Topic Key

Revision Topics	
Speed	1
Density	2
Pythagoras	3
Sequences	4
Solving equations	5
Constructions	6
Standard form	7
Area	8
Surface area	9
Rearranging formulae	10
Equation of a line	11
Drawing Graphs	12
Trigonometry	13

New Topics	
Solving simultaneous- no multiplying	14
Solving simultaneous- multiplying	15
Solving simultaneous equations graphically	16
Solving simultaneous by substitution	17
Basic probability	18
Draw tree diagram	19
Use tree diagram	20

Table *2*: Intervention Plan. The proposed topic plan for the intervention group. The topics are labelled in the table with a number such as 15, where the corresponding topic can be found in the Topic Key.

The interleaved topics seen in the intervention plan were all topics that the pupils had been taught, by me, previously this year. The use of this type of interleaving, using superficially dissimilar problems, was decided due to the nature of the current topic needing to be taught. The teaching of solving

simultaneous equations could be taught in a way that broke down various strategies (e.g. eliminations with/without multiplying, graphically). Simultaneous equations would have fitted well with superficially similar interleaving, whereby problems would look superficially similar, but learners would need to identify whether to multiply or whether they needed to add or subtract the equations. However, the complexity of the topic (entirely new to this year group) meant that teaching all strategies and having time for interleaved practice was not feasible for this topic in the allocated 50-minute lesson slot.

Therefore, given the positive results Foster et al. (2019) found in their study for both similar and dissimilar problem interleaving, I decided to include both. Initial interleaved practice was of dissimilar topics. Once each strategy of simultaneous equations was taught (and block practiced with four questions), a question of this type was included in the next lesson. This happened on a cumulative basis. Using Lesson 7 as an example, this lesson began with basic probability being taught, with four questions for practice. This was followed by 12 questions, including four solving simultaneous equations (one using elimination with no multiplying, one using elimination with multiplying, one graphically and one by substitution) which were dispersed between the other topics (e.g. Standard Form, Pythagoras).

The intervention was carried out over a three-week period. Which, for this year group was ten 50-minute lessons. The methodology for measuring the outcome of this intervention is outlined below.

3.5 How to measure the outcome of the intervention?

Using mixed methodology, containing both quantitative and qualitative data, enabled a wider understanding of both perceptions and attainment gains of interleaved practice. Mixed methodology allows for a multi-viewpoint and analysis. To improve triangulation of the data, I decided to use both questionnaires and an interview (Cohen et al., 2007), which also helps to establish a greater consideration of the issue. This mixed method approach was even more so necessary, due to the change in circumstances (COVID-19) midway through, to ensure completeness and allow for data to

be mutually corroborated (Bryman, 2006). With a need for a complete understanding of the issue and using pragmatism as an umbrella philosophy, a convergent design was followed with equal emphasis on qualitative and quantitative data.

3.5.2 Data collection

An overview of the data collected relative to the research question is detailed here:

Research Question	**Data collected and analysed**
1. What effects did the use of interleaved practice in lessons have on test results?	Scores attained in an assessment taken at the end of the intervention by both groups. Data analysed using an unpaired t-test to establish whether there was a significant difference between these. Question by question analysis comparing the proportion of pupils who answered correctly in each class
2. What were pupil perceptions of interleaved practice in lessons?	A survey of the intervention class (n=19) followed by group interview of 3 pupils. Axial coding used to break down analysis of data into key concepts.
3. How can interleaved practice be used in remote learning?	A survey of intervention class (n=19). Axial coding used to break down analysis of data into key concepts.

Table 3: An overview of the research questions and which data was collected and analysed in answer to those questions.

3.5.4 How to measure the effect the intervention had on test results?

As this question is focused on test scores, this data collection needed to be quantitative. The assessment was taken by both groups at the end of the intervention. The assessment was summative and based on topics covered previously in the scheme of work to include the newly taught simultaneous equations and probability. A copy of the assessment can be found at Appendix B.

Ideally a paired t-test would have been carried out to better measure the effects of the intervention for those pupils involved, by comparing their data before and after the intervention, however, given the naturalistic setting of a school this was not possible. Still, desiring a comparison of the two class

results and to ascertain whether interleaved practice has resulted in a statistically significant difference in attainment, an unpaired t-test was performed on the results.

In recognition of the limitations of using unpaired samples, alongside the t-test was a separate analysis, question by question, to identify any trends in performance in finer detail. This provided an opportunity to be able to identify whether there were any topics or types of questions which had been impacted more.

3.5.5 How to measure pupil perceptions of interleaved practice in lessons?

To assess pupil perceptions of using interleaved practice in the lessons, a questionnaire and group interview were used. As part of the questionnaire and interview, the term "interleaved practice" was used. Therefore, it was important to introduce this term to the class. This was done the lesson before the intervention began, where the class were shown definitions and examples to help with understanding. To further solidify this, during the intervention these definitions and examples were printed and stuck into the front of their 'interleaved books'. The methodology for the survey and interview will now be detailed.

Survey

Pupils of the intervention class were sent a brief questionnaire following the intervention of interleaved practice in the lesson. The purpose of surveying the intervention group was to gain a better understanding of student's perception of interleaved practice. The questionnaire was kept brief to allow for the pupils to complete this in the lesson to encourage high response rate. This was also to compensate for timing restrictions placed on the study because of the COVID-19 closures.

The questionnaire was sent to the pupils online in the lesson, they were encouraged to write as detailed and honest responses as possible. The questionnaire asked two questions:

1. What were the positives of using interleaved practice in lessons?
2. What were the negatives of using interleaved practice in lessons?

Responses were received immediately and were coded using axial coding (Cohen et al., 2007)

Given the brief nature of the questionnaire I wanted to further triangulate my data with an interview. The methodology for which is discussed below.

Interview

During this phase of the of the plan, the School setting was starting to become unsettled as talk of School closure began. Given the time constraints and imminent closure of the School due to the COVID-19 pandemic, the interview schedule was shortened and changed.

The original plan was to conduct one-to-one interviews to expand upon questionnaire answers and discuss the test. Whilst the timeline then made this unfeasible, a group interview was conducted with three pupils. Pupils were asked to volunteer to take part in a group interview.

The interview was semi-structured, which allowed for a two-way conversation to be had and to provide opportunity for further probing where necessary (Cohen et al., 2007), with the following key questions:

1. How are you finding interleaved practice?
2. What effect has interleaved practice had on your learning of simultaneous equations?
3. Thinking about the balance of new topic and interleaved practice, in your opinion, what is the right balance?

These questions were given to the pupils and they were given some time to consider their thoughts on them. I then asked each of them to respond in turn. The purpose of giving the participants the questions was to remove some of the influence from it being a group interview. It also allowed the interview to flow more, based on those key themes and any comments that each of them made, which then encouraged alternative or corroborative responses. Whilst the group interview will have resulted in the participants potentially being influenced by each other's ideas in some way, the benefits of bouncing ideas off each other allowed for further detail and clarification of answers. The interview was recorded using voice recording on a device, then written up verbatim afterwards. Following this, the data was then coded, and that coding discussed and agreed with the Deputy Headmaster.

3.5.6 How to measure how interleaved practice can be used in remote lessons?

Survey

As an adaptation to the evolving situation with the shutdown of the School, I wanted to gain a better understanding of what could be learnt from the first phase of the intervention and how this could influence remote lessons. Teaching of the full timetable was continued during the shutdown of the School, albeit remotely through Microsoft Teams. This provided the opportunity to inform pupils of the aim of the next stage and to send out a questionnaire to pupils online. As the pupils were spending a lot of time during lockdown on their devices in lessons, the questionnaire needed to be short. Therefore, it contained only two questions:

1. Do you want to continue with interleaved practice in your remote lessons?
2. How do you suggest interleaved practice be used in the remote lessons?

These two questions allowed me initially to ascertain whether pupil perceptions of interleaved practice was positive enough to want to continue it, and also to gain a better understanding of ways the pupils would benefit from interleaved practice being incorporated into the lessons.

3.6 Limitations

It must first be acknowledged the limitations of comparing two different classes for the quantitative data collection: two different samples and two different teachers. As described previously, the two classes' baseline data was compared, using previous department assessment data, and was indicative of largely similar groups. The sets being taught by different teachers could not be altered; in the department we each teach only one set of the same year group, at the same time. This difference of teachers has implications for the pupils: different teacher experience, style and agency will be experienced by the pupils in the two sets. However, to try to limit the implications of this on the study, the other teacher and I met and established a plan for both groups. Including an agreement that the other teacher would not use interleaved practice in the lesson (this was not something the other teacher usually did and, therefore, did not take away from pupils current learning). We also

established a list of topics that would be covered in the assessment and ensured we were in agreement that each of these topics had been taught previously to assessment standard and in accordance with the scheme of work the department was currently following. As described earlier, we also agreed on the strategies and methods to be taught for the new topics (simultaneous equations and probability).

Secondly, the limitations of the questionnaires and interviews should be addressed. During both the questionnaire and interview the term “interleaved practice” was used. Despite introducing and defining this term to the pupils, it should be recognised that it is possible for pupils to have misinterpreted this or taken their own interpretation of the phrase, thus reducing the validity of their responses to questions that involve the term “interleaved practice”. Additionally, the responses in both questionnaires and the interview are subject to bias. The validity of the data relies on good interviewing and whilst bias was attempted to be removed by not asking leading questions and remaining neutral in my response, this could have been affected by my inexperience in research interviewing. There are also the implications from the intervention and interviewing being carried out by myself. Being those pupil’s teacher may have caused the pupils to feel they needed to please. Indeed, during the interview one pupil said: “sorry Miss if that wasn’t what you wanted to hear”. This could be interpreted as them answering honestly or it could be interpreted as the pupil trying to please but apologising in case they were not successful. To try to limit the bias, pupils were actively encouraged to be as open and honest as possible throughout the process.

Acknowledging the limitations of this intervention and accepting of the implications of carrying out this intervention in naturalistic settings, affects what conclusions can be drawn for a wider field. Whilst it means that no concrete contributions can be made to this field of work, it does allow for considerations of the possible implications for this setting and age group.

3.7 Ethical considerations

The study had multiple participants, both pupils and teachers, and it is paramount that their interests were considered throughout. Therefore, ethical guidelines laid out in BERA (2018) were followed throughout. Importantly, the dignity and worth of participants were respected throughout the research. All participants were fully informed of the study and given the option to withdraw their data from the study.

3.7.1 Pupils

This intervention took place in my lessons and all pupils in the class took part in the project. The pupils in the parallel set were taught as normal by their classroom teacher and did not receive an intervention, however, they were fully informed of the study and the comparison being made between the two sets. In consideration of whether any participant would be negatively impacted by the study, both my set and the parallel set were receiving practice of topics following the scheme of work throughout the research period. Therefore, no group had a reduction in time spent being taught necessary topics and thus were not disadvantaged as a result of participating in the project. However, to further ensure this, following the intervention and test, the intervention class were given the opportunity to have an additional two lessons to secure their knowledge on simultaneous equations and probability. The parallel set teacher was also given all the interleaved resources used in the study to use with their class after the post-intervention test.

Participants for the interview were chosen by volunteering their names for interview. Before the interview, pupils were informed of the audio recording and verbal consent was given for this to happen. Pupils were given the option to decline interview with audio recording, without the expectation of a reason. The pupils were informed that the results would be used in the research project and would not be used for re-setting or parental communication. Given the contextual setting, this was to reduce examination anxiety and anxiety from potential parental pressures.

3.7.2 Teachers (including the author)

The other teacher of the parallel set and I collaborated on the project and discussed the outcomes. Whilst unannounced tests fall under the School's normal procedure and policies and is common practice in the School, after collaborating with the teacher of the parallel set it was decided to inform the pupils to expect the test on the date it was given.

It is important to recognise that, this being classroom-based research, I am the main teacher involved in the research and the implementation of the intervention. As a result of my participation, bias cannot be eliminated but using awareness and constant reflection I remained as objective as possible (Menter et al., 2011). Alongside my objectiveness in the project, there must be a recognition of the effects that being the participants' teacher will have had on the study. Whilst I encouraged pupils to be open and honest in their responses, it must be acknowledged that the power asymmetry may have impacted pupil responses. To potentially remove this bias another teacher could have carried out the interviews, however time restrictions (given the COVID-19 situation) did not allow for this. Similarly my role as teacher and researcher may have impacted their decision to participate in the study, and to reduce this, pupils were reminded throughout the study that they had the option to verbally opt-out of their data being used in the study.

3.7.3 Data

Data gathered from the test was recorded anonymised, and accurately and non-objectively reported. Confidentiality was addressed using data cleaning, and the write up of the project included no real names or identifiable information. Finally, all data collection and gathering of assessment results falls within normal practice and policies of the school.

4. Findings and Discussion

The findings from the data gathered will now be discussed.

4.1 What effects did the use of interleaved practice in lessons have on test results?

Pupil test sheets were collected in and marked by their set teacher, then cross marked by the other teacher. It is important to note here that both myself and the other teacher identified the test, that the pupils were given, as more challenging than one we would perhaps have set. However, removing the bias from the assessment was important for the validity of the results. We both felt that our respective classes would find the test challenging. The individual and class results are demonstrated in Table 4 below where Class A is the intervention class and Class B in the non-intervention class.

	Pupil Marks	
	Class A	Class B
	20	12
	18	16
	17	11
	13	16
	16	14
	20	12
	14	9
	13	22
	25	12
	36	13
	32	15
	14	4
	34	31
	16	
	19	
Mean	20.5	14.4
St.Dev	7.7	6.5

Table 4: Test Results. Individual pupil results from the test taken by both groups, class mean and standard deviation.

Due to pupil absence or pupils leaving the School, the sample size declined before the test phase (Class A, n=15; Class B, n=13), although these were still large enough to be able to draw some comparisons and complete the statistical analysis. The data in Table 4 demonstrates that, on average, Class A performed better than Class B (Class A $\bar{x} = 20.5$, Class B $\bar{x} = 14.4$). To understand whether there was any statistical merit to this, an unpaired t-test was performed. Prior to this, an f-test was performed to determine if the variances for Class A and Class B (59.6 and 42.3 respectively) were equal. Using

Using $H_0: \sigma_1^2 = \sigma_2^2$ and $H_1: \sigma_1^2 \neq \sigma_2^2$, the f-test indicated that the two sample variances were equal, at the 5% level of significance (P(F≤f) = 0.2788). Thus, enabling an unpaired t-test to be carried out with assumed equal variances. The results of the t-test was as follows:

t-Test: Two-Sample Assuming Equal Variances

	Class A	*Class B*
Mean	20.5	14.4
Variance	59.6	42.3
t Stat	2.24	
P(T<=t) two-tail	0.0342	
t Critical two-tail	2.06	

$H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 - \mu_2 \neq 0$
0.0342<0.05
Reject Ho

The 2-tailed t-test performed on the two class samples was significant (P(T≤t) = 0.0342), at a 5% level of significance. Indicating that the observed difference between the sample means (20.5-14.4) is convincing enough to say that the average result from the test differs significantly. These results demonstrate a significant difference in the two classes, enabling a successful answering of the study's first research question: "What effects did the use of interleaved practice in lessons have on test results?". The intervention, as shown by the above statistical analysis, has shown that the interleaved practice has a statistically significant positive impact on test results.

Additionally, a more in-depth analysis was undertaken on the results, question by question (raw data – Appendix C), to ascertain whether there were any noticeable findings for each of the questions. The findings from this analysis are demonstrated in Figure 3, which shows the proportional difference between the two classes. Class A and Class B had unequal sample sizes and to account for this in the analysis the proportion of pupils in each class was calculated, and compared, rather than the number of pupils. The analysis identified several topics whereby Class A, the intervention class, had a positive difference in the proportion of pupils who answered the question correctly. In only one topic

(rearranging formulae) did Class B have a higher proportion of pupils answering correctly. In six topics there was no significant difference between the proportion of correct responses: Pythagoras, sequences, constructions, area, solving equations and surface area. In Class A one pupil accounts for 6.7% of the class and in Class B one pupil accounts for 7.7% of the class. Therefore, no significant difference (*d%*), as mentioned above, between the proportions was deemed to be $-8 \leq d \leq 8$, as this would be a difference equivalent of one pupil, or less.

Figure 3: The percentage of pupils who answered each question (by topic) correctly by class.

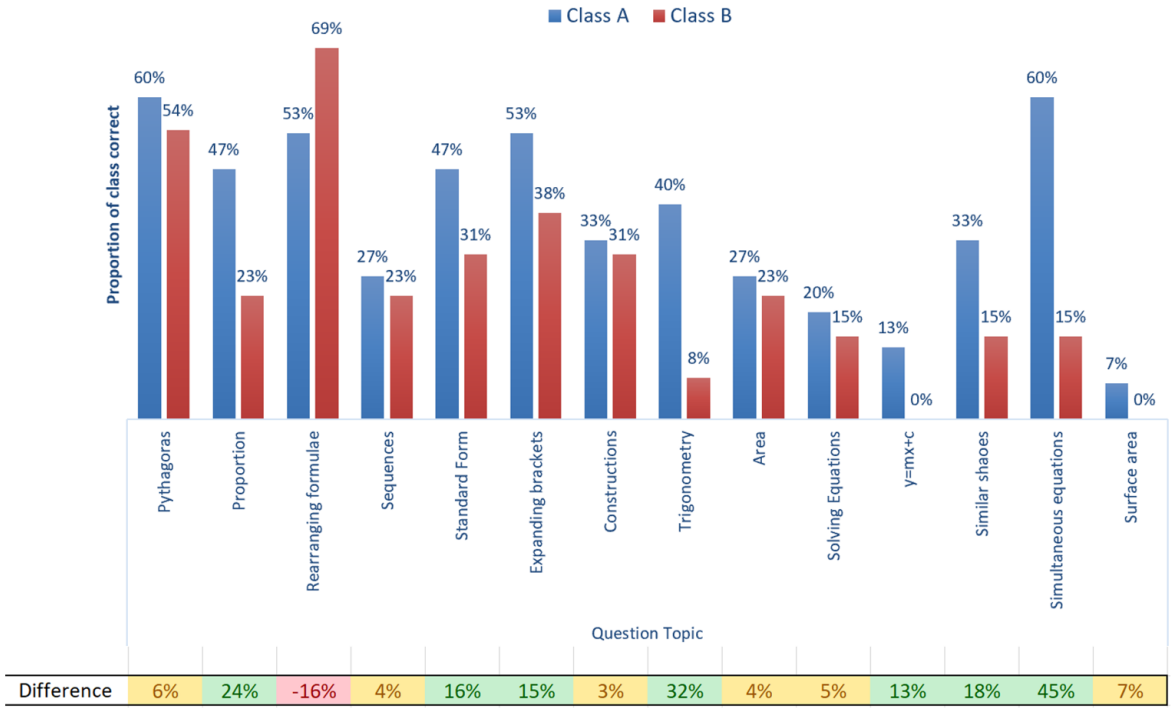


Figure *3:* Assessment analysis. A comparison between the proportion of pupils who answered each question (listed by the topic) correctly. The percentages shown are the percentage of pupils, from the respective class, who correctly answered that question in the assessment; whereby correctly answering means gaining full marks for the question. Pupils who only achieved part marks in the question were not counted as correct. The difference between the classes ($‰_a$-$‰_b$) is green if class A, the intervention class, had a higher proportion who correctly answered that question, red if lower, and orange if no significant difference.

However, significantly, on seven different topics a higher proportion of pupils in the intervention class were able to achieve full marks than in the non-intervention class. These topics were: proportion, standard form, expanding brackets, trigonometry, y=mx+c, similar shapes, and simultaneous equations. The most noticeable differences were in proportion (24%), trigonometry (32%), and

simultaneous equations (45%). It is this latter difference that is most significant, 60% of the intervention class were able to answer the simultaneous equations question compared to only 15% of the parallel set. This is a significant finding because simultaneous equations was the new topic taught throughout the beginning of the intervention, and subsequently practised in an interleaved manner. Indeed, this analysis is indicative that learning simultaneous equations with interleaved practice has produced superior results.

This finding demonstrates that not only has the intervention had a positive effect on test results overall, meaning a greater recall of previous topics and a positive effect on the learning of the current topic. These findings of superior results for interleaved practice is comparable to findings from previous studies (Taylor & Rohrer, 2010; Rohrer et al., 2015). Whilst the quantitative data does not provide the opportunity to understand why that is, the literature may provide some suggestions. Superior results in the test may have been because of improved ability to discriminate between problems (e.g. Rohrer et al., 2015; Foster et al., 2019). Specifically, these findings are further support of positive effects found with superficially dissimilar problems and the ability to assign the correct strategy (Rohrer et al., 2014; 2015) as the test used mixed topics. Other reasons for superior test results found in the data could be due to greater long-term retention from retrieval opportunities (Rawson & Dunlosky, 2012; Adescope et al., 2017), or spaced study (Schutte et al., 2015; Peterson-Brown et al., 2019).

4.2 What were pupil perceptions of interleaved practice in lessons?

As discussed earlier in the methodology, to gain an understanding of pupil perception of interleaved practice in lessons all pupils in the intervention class were surveyed, and a small group were interviewed. Firstly, the results of the questionnaire will be discussed, followed by discussion of the group interview.

4.2.1 The Questionnaire

The questionnaire was sent online to all pupils in the intervention class. With answers to: 'What are the positives of interleaved practice in lessons?' and "What are the negatives of interleaved practice in lessons?" coded into three categories: content, structure and understanding. Each of which will be discussed below.

Content

A common theme occurring in the data, relating to content of the interleaved practice, was the aspect of the inclusion of previously leant topics. When asked about what the positives of interleaved practice comments such as "I like the revising aspect - going over things we have done before" and "It ensures you don't forget topics you've already learned". Their recognition of this, may come from completing questions and getting better each time or it might be from recognising they have forgotten certain topics and skills. As suggested by Rohrer et al., (2015) the interleaved practice has provided an opportunity to practise the skill they are expected to learn, or to learn that they cannot do it.

Another reoccurring theme in comments made by pupils in response to the positives of using interleaved practice was based on variety. The following comments were made by individual pupils in their response:

> "It had lots of different types of questions, easy and hard."
> "Helps all subjects"
> "It's a mix of questions"
> "You practice lots of different topics"
> "I really enjoyed the fact that it was a change each lesson."

The pupils here, identified the variety of topics and mixture of questions to be a positive of interleaved practice. The first comment identified the variety in difficulty of questions to be a positive also. The final comment expressed joy at the change each lesson. These comments suggest greater engagement and enjoyment from the lessons when they involve this mixture of questions as part of the interleaved

practice. Similarly, in previous studies (Wahlheim et al., 2011; Metcalfe & Xu, 2016) perceived trivial tasks led to reduced engagement and study time. Which suggests, the inclusion of challenging questions and mixed topics included in the design of this interleaved practice was beneficial to learning.

This analysis is also indicative of support towards using the remote interleaved questions (Foster et al., 2019) as this requires mixed topics to be included in the interleaved practice rather than the superficially similar questions introduced in other studies (Rohrer &Taylor, 2007; Taylor & Rohrer, 2010). As this provision of mixed topics problems was perceived as one of the benefits of interleaving.

Structure

An important element of the practice that was highlighted in pupil responses was the structure. Responses made by different pupils pertaining to this included “It has a good structure”, and “kept everything going at a fast pace” and “it’s a small test so breaks it up”. The pupils here see the benefit of the structure of the interleaved practice to be that it breaks up the lesson and ensures that it continues at a fast pace. This change in activity and perhaps the movement onto the next activity provides motivation and improves engagement. This is similar to what Metcalfe & Xu (2016) experienced with learners maintaining engagement over time when not receiving blocked practice.

Whilst a clear common positive was felt by the structure of the interleaving, not all comments around the structure were positive. A negative of interleaved practice, identified in the qualitative data associated with structure, was the repetitive nature of it as indicated by these comments made by individual pupils:

> “It's quite repetitive.”
> “Too many of the same questions.”
> “Repetitive Questions.”
> “It's quite repetitive, lesson after lesson”

Pupils here are identifying a dislike of the repetition. Interestingly there is mention of there being “too many of the same questions”, yet throughout the course of the intervention no two questions were

the same. Therefore, the pupil here perhaps meant the repetition of topics over the course of the interleaved practice. Another repetitive factor here is that this structure of the duration of the intervention: the structure of each lesson was the same. Pupils may have found this lesson structure the repetitive factor.

Timing is another important element of the structure of the lesson that includes interleaved practice. In a classroom setting the practice is limited to the allocated lesson time, in this case 50 minutes. As explained earlier in the methodology the design of the lesson, being 4 questions new topic and the rest interleaved, was based on the advice of Rohrer et al. (2014). Therefore, it was important to gain pupil perceptions on the design of this and whether the pupils perceived the balance to be right between current topic practice and interleaved practice. When asked "do you think the balance was right between new topic practice and interleaved practice" 74% of the pupils answered yes. This left 26% of the pupils expressing the balance was not right.

The comments made by individual pupils associated with timing perhaps offer an explanation as to why that is:

> "Sometimes you need a bit longer on the tasks."
> "It takes up more of the lesson than what we're actually learning."
> "Takes up half the lesson."
> "We only get a short time on the current topic."

Each comment above suggests that these pupils perceive that a negative of the structure of the practice was that not enough time is spent on the current topic, implying that they would rather have more time spent practicing the new topic than the interleaved practice. However, these comments only came from four pupils and needed further exploration in the interviews.

Understanding

The two previous sections demonstrate important findings were made from pupil perceptions of the structure and content in the delivery of the interleaved practice. Another key set of perceptions

centred around the learning the pupils felt was achieved using interleaved practice. A number of pupils mentioned the improvement of understanding, such as "you become more confident with some topics", "I think I'm better now at the topics", and "I started to feel like I can understand other bits we did before". These comments demonstrate perceptions that the interleaved practice had improved their confidence and improved their understanding of previous topics. Whilst these perceptions have rarely been explored in the literature, this improvement in knowledge has been found in the increased test scores (e.g. Mayfield & Chase, 2002; Taylor & Rohrer, 2010; Foster et al., 2019). It might be suggested that pupils have recognised their improvement in answering the questions correctly, resulting in a perception of improvement of understanding of topics.

Associated with this was the aspect of the opportunity for revision and retrieval of knowledge. Comments of such nature, made by individual pupils were:

> "It helps you refresh your memory"
>
> "That we are going over subjects keeping them fresh in our memory. They help with end of year exams"
>
> "Helps me remember stuff we have learnt in the past and get better at the stuff we learn in class"
>
> "… refreshes my memory on previously covered topics."
>
> "It helps you remember subjects because you always have to think about them"

Pupils here saw the positives of the revision aspect of interleaved practice. The retrieval element in the process of completing the questions is a highlight, leading to the pupils perceiving this as a way of keeping "fresh" things in their memory. The integral nature of retrieval in interleaved practice, as suggested by the literature (Rohrer et al., 2020b), is providing the pupils with the opportunity to retrieve and strengthen the knowledge

As was found in the literature, the notion of desirable difficulty highlighted that the learning process involved might not be experienced as a positive one by the pupils (Carpenter et al., 2008; 2012; Dunlosky et al., 2013; Rohrer et al., 2015). This was certainly evidenced in the responses made by

pupils when answering what the negatives of interleaved practice were. Comments, made by individual pupils, suggesting the difficulty they experienced during practice were:

> "Some of the questions aren't clear or we haven't really learnt the topic."
> "It can sometimes be confusing as it jumps from topic to topic."
> "Some of them are not like what we have done before."
> "Some questions are difficult."

There are multiple comments here highlighting that pupils found the process of switching between topics, and being able to identify the correct strategy, difficult. These pupils found the questions difficult to complete. This experience of difficulty experienced by the pupils countered by positive test results is a demonstration of desirable difficulty (Carpenter et al., 2008; 2012; Dunlosky et al., 2013; Rohrer et al., 2015). Whilst the process may have felt uncomfortable or challenging at the time, when coupled with the improved test results (compared to the parallel class) this provides evidence of the positive effects of this.

One pupil mentioned that the questions were difficult to complete in the time given: "sometimes there were some questions in the sheets that would not be completed in one lesson and build up over time", and this left incomplete questions building over time. This may suggest that the methodology needs improving and that the interleaved practice worksheets were too ambitious in the amount of questions they contained. The time-consuming nature of the interleaved practice was also experienced by participants in Rohrer et al.'s (2020b) study, suggesting that if teachers are to implement interleaving within the lesson appropriate time allocation should be made for this.

4.2.2 The Interview

A group interview was conducted with three pupils in the intervention class.

Once the interview was conducted the transcript was coded to look for key concepts discussed or mentioned. The key concepts that arose in the analysis of the interview transcript were: revision,

confidence, understanding, difficulty, and structure. Each of which will be discussed below. An extract of the coded transcript can be found in Appendix D.

Revision

The two pupils here are expressing a positive impact of interleaved practice in that it provides an opportunity to review previously learnt topics.

> "I think it's really good to revise the topics that we have already done" Pupil A
>
> "I find it good to revise the topics that we get like stuff we go over" Pupil C

Pupils are using the practice time as a memory recall tool. The element of retrieval (Adescope et al., 2017) here is one which the pupils themselves recognise the benefit of as a part of the interleaved practice. This also coincides with the findings from the questionnaire, that pupils find the revision of previously learnt topics a positive of interleaved practice.

Confidence

An identification of a theme of confidence and awareness of ability within taught concepts was made from the analysis. When asked whether they had managed to answer the question on simultaneous equations that appeared in the interleaved practice from that day and how that made them feel.

> "… actually, I knew what I was doing because I'd learnt it. So, like today in the interleaved I got it right, I was kind of pleased with that." Pupil B
>
> "It felt quite good (laughs) to be honest I hadn't done well on the questions before, but today … I sort of remembered how to do it so it was good yeah (laughs) It was a relief." Pupil A

Both pupils here were visibly pleased when asked these questions, they were either laughing or grinning. Interestingly Pupil A had mentioned earlier in the interview his dislike of the interleaved practice due to feeling like he "wasn't getting it" before we moved on. However, he expresses here his satisfaction at managing the simultaneous equation question in that lesson. The pupil mentioning his "relief" was quite a significant comment to make. Clearly this pupil had been lacking confidence in his ability before this point.

Interestingly something else that came to light in the interview was a lack of confidence compared to peers and how that might affect how they wish to complete practice and ask for help.

> "some people like move really far ahead of everyone else and it gets a bit funny." Pupil C
>
> "Before when it was a new topic, I just left the question because I was struggling with it. I didn't want to get it wrong in both the classwork and the interleaved because then it looks like I really didn't get it." Pupil A

Here Pupil C expresses his dislike of feeling as though the class are not moving at the same pace. Pupil A here lacks confidence in his ability to such an extent that it is preventing him from even trying the question for fear of getting it wrong. Despite the worksheet of the interleaved practice not being collected in or seen, he did not want it to look like he got it wrong. Yet this is the pupil who above was quoted expressing relief at being able to answer the questions. Clearly there is a feeling of difficulty experienced in the practice phase, like previous studies who experienced desirable difficulties (Carpenter et al., 2008; 2012; Dunlosky et al., 2013; Rohrer et al., 2015) However, the strength of the lack of confidence amongst peers has a role to play here, affecting the pupil's engagement in the interleaved practice. This is a factor that needs considering in the implementation of interleaved practice.

Understanding

As with the questionnaire, the understanding and learning experienced during the intervention was another common theme found in the interview. Pupil A expressed a negative of the interleaved practice being that he was struggling with understanding simultaneous equations and felt that he needed longer practice of these to secure his knowledge:

> "with simultaneous equations I didn't get it, so probably like doing it for the entire lesson for 50 minutes would probably be more helpful, just do it do it do it. Whereas doing four questions might not have helped" Pupil A

The pupil expressing his desire to "do it do it do it" shows that he would have found it to be more helpful to have more than the allocated four questions on the strategy. This raises the question about overlearning discussed in the literature and whilst there may not have been any effect found on test results when overlearning (Rohrer & Taylor, 2006; Rohrer et al., 2015), this overlearning strategy appears preferable to the pupil.

During a further probe of this response, when asked if in the last lesson Pupil C had managed to get the simultaneous equation question (embedded within the interleaved practice) right, they responded with:

> "No, I did like after you told me it was wrong, but at first I got it really out. And I think it's just like I know it but I just forget … I think it makes it more confusing when you've done something else and like lots of things" Pupil B

Firstly, the pupil referring to being told he was wrong and then self-correcting, demonstrates a positive impact of pupils being able to review answers and make corrections during the interleaved time, a factor which needs considering when implementing interleaving. Secondly, the pupil identifies that there is a difficultly when faced with the need to retrieve information and failing to do so. It would be the hope that as this continued, their retrieval of material would strengthen and the "forgetting" that the pupil mentions would improve (see Adescope et al., 2017). This act of lacking the ability to retrieve the information at times was also echoed by Pupil C: "I just forgot what you were meant to do." The ability to select the appropriate strategy, which has been argued is more difficult than the execution (Rohrer et al., 2015), was difficult for these pupils. Which poses the question: are the four questions enough, for understanding but also confidence?

Difficulty

Alongside some concern over the difficulty of remembering how to do the current topic there was an acknowledgement that the pupils found completing the interleaved practice difficult.

> "and then doing other things I don't know makes my brain hurt a bit" Pupil A
>
> "Yeah I kind of agree, you do something like simultaneous equations and then perpendicular bisector and they're just like really different" Pupil B

Both pupils here expressed a perception of difficulty in switching between content. They acknowledged that, for them, dealing with a switch in question content was a challenge. Whilst this may have produced desirable difficulty (Carpenter et al., 2008; 2012; Dunlosky et al., 2013; Rohrer et al., 2015), it may suggest a preference for the interleaved practice to contain superficially similar questions.

When asked to comment on the balance between interleaved practice and current topic practice, Pupil A responded with:

> "I think the amount of questions in the interleaved and simultaneous equation is fine, but I think trying to do simultaneous equations the next day is hard." Pupil A

This is particularly interesting given that this was the same pupil who, before, had said that they wanted more practice of current topic. Which leads to the question as to whether the pupil here wanted more practice to overcome that fear when facing it the next day? Pupil A also mentions here that they found completing the current topic the next day hard.

Structure

As mentioned above, the pupils were asked whether the balance between new topic and interleaved was right and if not, how they would rather have it. The following responses relate to the structure of interleaved practice:

> "I think we should do definitely more questions of the thing we're actually learning" Pupil B

> "I would probably do like ten questions we have been set for that topic and then do the interleaved. Your brain then remembers more of the work you are doing in the class, so you get more chance of remembering it more the next lesson." Pupil A

Going forward, both pupils would like to have more questions on the current topic (compared to the four in this intervention). The pupils here both clearly perceive the act of completing more questions on the current topic, in this case simultaneous equations, to result in more secure knowledge and memory of it. However, given Rohrer and Taylor's (2006) (also see Rohrer et al., 2015) finding on overlearning, more secure knowledge may not be the resultant of more questions. But perhaps there ought to be recognition here of the desire from these pupils to feel more confident in the subject.

Alongside the above, there was recognition by the pupils of the desire for autonomy:

> "If I'm finding it hard, I want to be able to do more. I want the choice of moving on if I feel confident but staying if I don't" Pupil B

> "Yeah I think the interleaved practice should be shorter … because sometimes we don't even get through it … and more flexible timings if you're a bit stuck on the questions and get help. " Pupil B

> "I would rather do more questions like longer like we would in a normal lesson in the lesson and then once you are finished go on to the interleaved practice." Pupil C

The theme here is the desire to be able to have control of when they feel ready to move on from the current topic practice to the interleaved practice. This might work well for differentiation within the lesson but would be reliant on pupils being able to self-assess effectively.

4.3 How can interleaved practice be used in remote learning?

The first question addressed to the pupils in the online survey was whether they wanted interleaved practice to be included in the remote lessons following the closure of schools (see COVID-19 impact). All pupils responded yes, providing strong evidence of a positive perception of the use of interleaved practice.

Following this, pupils were asked for suggestions for ways interleaved practice could be used in remote lessons. Again, the responses to these were analysed and coded.

Some pupils wanted the element of choice to be included: "maybe if you've finished what we were doing go on to the interleaved practice" and "if we complete work that you have set us then we do the interleaved when we have finished the set questions." Both pupils mention here that they would like to move onto the interleaved practice once they have completed the work.

Some pupils wanted the interleaved practice to follow the same format as before, with comments such as "maybe do the interleaved practice in the second half of a lesson", "do all practices in 2nd half of lesson" and "we do a bit of the actual lesson then finish with an interleaved". These pupils all mention the desire for the lesson to be finished with the interleaved practice.

Similarly, one comment suggested a similar format only with a review suggestion:

> "Much like in the classes we could do the new topic questions for the first bit then upload a sheet of interleave questions to one note and then perhaps for the questions which are answered wrong by a majority of the class can be practiced at the beginning along with the new topic."

This pupil is highlighting a desire for there to be review of wrongly answered questions the following lesson alongside the new topic.

Finally, one comment pertains to the variability of including interleaved practice: "I think we should not do it EVERY lesson but maybe twice a week. Some lesson yes and some lesson not." This suggests

a desire for variety, like those pupils who mentioned in the questionnaire that repetition was a perceived negative of interleaved practice.

4.4 Limitations

Whilst the data gathered in this project is positive towards the use of interleaved practice, it is important to acknowledge the limitations which may affect the reliability of this data.

Firstly, the difficulty of the test may have resulted in pupils losing confidence and therefore underperforming; the reduced amount of correct answers as the test progressed could also be indicative of this. As the pupils lost confidence, they answered more questions incorrectly. The test was also limited to a time frame and all skills of new topics were not tested. For example, the test only contained the solving of simultaneous equations by elimination (with multiplying). Therefore, not all skills taught were tested.

Secondly, whilst the data demonstrates a statistically significant result on the test results of the class, this may not hold true for everyone. A paired t-test would have needed to be carried out to explore whether there was a significant impact for the same sample. Data pre and post intervention would have also enabled analysis on an individual basis. However, as previously explained, the naturalistic setting of this study did not allow for this.

Thirdly, it should also be recognised the time limitations placed on the intervention and the impact this will have had. The intervention was only able to be carried out for two weeks with data collected immediately after this. Therefore, any conclusions made cannot be done so on a long-term retention basis, as this was not able to be measured.

Finally, the sample chosen for the intervention was small and restricted. It was carried out with one class, in one-year group. Therefore, conclusions made should be on this basis and thus reduces validity in applying these conclusions outside of this sample. However, it does provide evidence to extend the research.

5. Conclusions and Implications

5.1 Conclusion

This research aimed to determine whether interleaved practice is beneficial in mathematics, to improve outcomes in the classroom. In this final summary of findings, I will focus on summarising the evidence found for each of the three research questions. Following this, is a consideration of the implications of these findings.

5.1.1 The effects interleaved practice had on test results

The effects the interleaved practice intervention had on test results were:

1. Superior test results for the class when compared with the parallel class who were taught using blocked practice.
2. A higher proportion of pupils were able to answer the question on the new topic being taught through interleaved practice (simultaneous equations).

In accordance with literature (Taylor & Rohrer, 2010; Rohrer et al. 2015; Rohrer et al., 2020b), findings from this research revealed higher test scores, with statistical significance, for those pupils experiencing interleaved practice on an immediate test; suggesting that interleaved practice strengthens the retention of knowledge of previously learnt topics and the ability to apply that knowledge in an examination setting.

Alongside strengthening of knowledge of previously learnt topics, interleaved practice proved beneficial to the learning of the new topic. This research demonstrated that a higher proportion of pupils in the interleaved group were able to answer the question on simultaneous equations in an examination condition.

Whilst this research does not provide the reasons as to why these outcomes were found, previous literature would suggest they may be as a result of interleaving strengthening problem discrimination and problem strategy association (e.g. Rohrer et al., 2015; Foster et al., 2019). Alongside this are

further learning strategies that are at the foundation of interleaved practice, particularly when used in naturalistic settings, spaced practice and retrieval. Spacing the presentation of problems improves learning (e.g. Schutte et al., 2015; Peterson-Brown et al., 2019). Retrieving knowledge, especially when presented with it when on the verge of forgetting, is also beneficial for learning (Rawson & Dunlosky, 2012; Adescope et al., 2017).

However, it must be acknowledged that these comparisons were made with a parallel set not a control group. As such, any results need to be recognised regarding their limitations from lack of control of other variables; significantly, the difference in teacher and the impact this will have had on the learning experience for pupils.

5.1.2 Pupil perceptions of interleaved practice

The key positives of using interleaved practice, from the perspective of the pupils, were:

1. The variety in questions that interleaved practice offers
2. The change in activity
3. The revision of previously learnt topics

In the class questionnaire, pupils revealed that they liked the variety that the interleaved practice offered. Similarly, the change in activity, from learning the new topic to then moving onto the interleaved practice the pupils faced each lesson was perceived as a positive and a way to keep the lesson moving “at a fast pace”. This suggests that this variety provides motivation for greater engagement in lessons, as found in previous studies (Wahlheim et al., 2011; Metcalfe & Xu, 2016).

Echoed in both the class questionnaire and the group interview, was the strong message that pupils found the revision aspect of completing interleaved practice a large positive. This research found that pupils recognise the benefit of being presented with previous topics to strengthen their knowledge of these.

The key negatives of interleaved practice, from the perspective of the pupils, were:

1. Repetitive
2. Difficult
3. Not enough time spent practising new skill

One of the drawbacks discovered in this research was that pupils found the repetitive nature of interleaved practice a negative. Experiencing the same lesson structure for ten lessons may have resulted in this opinion, and therefore teachers, who adopt interleaved practices, may want to vary the structure of teaching of the new topic and presentation of interleaved practice.

Similar to previous studies, where teachers said that they thought their students found interleaving harder and more time consuming (Rohrer et al., 2020b), this research demonstrated that the pupils, expressed themselves (rather than from teachers), found the practice phase challenging. Data from the interview revealed that they particularly found the remote interleaved practice (Foster et al., 2019), jumping between topics, hard. Therefore, the difficulty experienced during the practice phase alongside superior test results demonstrates desirable difficulties being achieved (Rohrer & Taylor, 2007; 2009; Rohrer et al., 2020b).

Timing was an issue for some, with pupils expressing that they needed longer on the current topic. This was echoed in the interview where the pupils expressed a desire for more questions on the current topic to be included in the lesson, or for the choice to move on when ready rather than only being presented with four questions. In the interview this was also found to be the case, with the pupils expressing a desire for more current topic questions. This disputes Rohrer et al.'s (2006) suggestion that pupils should be presented with no more than three or four questions of the same skill in succession. Whilst the test results were positive using only 4 presentations of the new skill, the qualitative data revealed that some pupils did not feel confident with the current topic. However, given the finding that a higher proportion of the intervention class were able to correctly answer the new topic question in the examination, this would suggest that the difficulty pupils faced with lacking confidence in the new topic paid off in the longer term, thus providing evidence to support the

productive failure theory (Kapur, 2008; 2014). A balance must be had between outcome and experience, pupil's performance in examinations is important but so is their confidence and engagement in the subject.

5.1.3 Use of interleaved practice in remote lessons

As this was an addition to the research, due to unprecedented times, this was a much shorter data collection. Nevertheless, some key findings were made from the data.

All pupils wanted interleaved practice to be continued into remote lessons, evidencing the positive experience pupils had of interleaved practice when it was used in normal lessons. A common theme in the questionnaire was that pupils wanted the interleaved to remain in the second half of the lessons. However, pupils expressed the desire for autonomy, they wanted to be able to decide for themselves when to move on to the interleaved practice. This may be because of the lack of confidence in the current teaching topic, as earlier discussed. Due to the unique nature of the relevance of this research question, it is not something that has been explored in previous literature.

5.2 Implications

There is strong evidence for the use of interleaved practice in mathematics, in both previous literature (e.g. Rohrer et al. 2014; 2015; Foster et al., 2019) and in this study. As a result, the research has been discussed with colleagues in the department, resulting in a reflection for improvement of professional practice.

To progress this research into an expansion within my own setting, the use of interleaved practice should be trialled and evaluated with other year groups. Additionally, it would be interesting to measure whether the same benefits are found with high or low attaining sets. In expanding this to other year groups and attainment sets, the feasibility for teachers must be considered. All interleaved resources created in this study were shared, with an agreement to form a collective set of resources to increase feasibility for teachers implementing interleaved practice in their own classrooms.

This study provides evidence towards a gap in the field of research on interleaved practice: the perceptions of the pupils. As this research was conducted using a small sample, with limited diversity, a larger scale study, in different social contexts, is required to establish further conclusions of statistical significance. Due to the limitations of the intervention, because of COVID-19, it would be beneficial to repeat the study scaling the time of the intervention and measuring the difference in delayed tests.

6. References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659-701.

Agarwal, P. K. (2019). Retrieval Practice & Bloom's Taxonomy: Do Students Need Fact Knowledge Before Higher Order Learning? *Journal of Educational Psychology*, *111*(2), pp. 189–209.

Agarwal, P. K., & Bain, P. M. (2019). *Powerful teaching: Unleash the science of learning*. SanFrancisco, CA: Jossey-Bass.

Agarwal, P. K., D'Antonio, L., Roediger, H. L., III, McDermott, K. B., & McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of Applied Research in Memory & Cognition, 3*, 131–139

Birnbaum, M. S., Kornell, N., Bjork, E. L., Bjork, R. A. Why interleaving enhances inductive learning: the roles of discrimination and retrieval. *Memory & Cognition, 41*(3), 392-402.

Bjork, R. (1979). Information-processing analysis of college teaching. *Educational Psychologist, 14*(1), 15-23.

Bjork, R. A. (1994). Memory and meta-memory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bjork, R. A. (2013). Desirable difficulties perspective on learning. In H. Pashler (Ed.), *Encyclopedia of the mind*. Thousand Oaks: Sage Reference.

British Educational Research Association (BERA) (2018) *Ethical guidelines for educational research*. 4th edn. Available at: https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018 (Accessed: 20 April 2020).

Bryman, A. (2006). Integrating Quantitative and Qualitative Research: How Is It Done? *Qualitative Research, 6*, 97-113

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*(2), 438–448

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review, 24*(3), 369–378.

Carvalho, P.F., & Goldstone, R.L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review, 22*, 281-288.

Carvalho, P. F., & Goldstone, R. L. (2019). When does interleaving practice improve learning? In J. Dunlosky & K. A. Rawson (Eds.), *Cambridge Handbook of Cognition and Education* (pp. 411–436). New York: Cambridge University Press.

Casey, D., Carroll, C., & Crowley, J. (2018). Rethinking the Mathematics Worksheet in Higher Education: Embracing the Application of Spaced and Interleaved Practice. *Irish Journal of Academic Practice, 7*(1), Article 2.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354-380.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing Effects in Learning: A Temporal Ridgeline of Optimal Retention. *Psychological Science*, *19*(11), 1095–1102.

Cohen, L., Morrison, K., & Manion, L. (2007). *Research Methods in Education (6th ed*.). Taylor & Francis Routledge.

Cull, W.L. (2000), Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14,* 215-235.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14(*1), 4-58

Firth, J., Rivers, I., & Boyle, J. (2019). A systematic review of interleaving as a concept learning strategy: A study protocol. *Social Science Protocols*, *1*, 1-7.

Foster, N., Mueller, M., Was, C., Rawson, K. and Dunlosky, J., (2019). Why does interleaving improve math learning? The contributions of discriminative contrast and distributed practice. *Memory & Cognition*, *47*(6), 1088-1101.

Kang, S. & Pashler, H. (2012). Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast. *Applied Cognitive Psychology. 26*. 97-103

Kang, S., Lindsey, R., Mozer, M., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review, 21*(6), 1544–1550

Kapler, I., Weston, T. & Wiseheart, M., (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. *Learning and Instruction*, *36*, 38-45.

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*, 379-424.

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, *38*, 1008-1022.

Karpicke, J. D. and Grimaldi, P. J. (2012), 'Retrieval-based learning: A perspective for enhancing meaningful learning', *Educational Psychology Review*, *24*(3), 401–418.

Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories: Is Spacing the "Enemy of Induction"? *Psychological Science*, *19*(6), 585–592.

Le Blanc, K., & Simon, D. (2008). *Mixed practice enhances retention and JOL accuracy for mathematical skills*. Chicago: Paper presented at the 49th Annual Meeting of the Psychonomic Society.

Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis*, *35*, 105-123.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology. 27.* 360-372

Menter, I., Elliot, D., Hulme, M., Lewin, J., & Lowden, K. (2011). *A guide to practitioner research in education*. Bodmin: SAGE Publications.

Metcalfe, J., & Xu, J. (2016). People mind wander more during massed than spaced inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 978-984.

Nazari, K. and Ebersbach, M., (2019). Distributed practice in mathematics: Recommendable especially for students on a medium performance level?. *Trends in Neuroscience and Education*, *17*, 100122.

Ostrow, K., Heffernan, N., Heffernan, C., & Peterson, Z. (2015). *Blocking vs. Interleaving: Examining Single-Session Effects Within Middle School Math Homework*. In International Conference on Artificial Intelligence In Education (pp. 338–347). Springer International Publishing.

Pashler, H. J., Cepeda, N. K., Carpenter, S., & Rohrer, D. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin and Review*, *14*(2), 187-193.

Petersen-Brown, S., Lundberg, A., Ray, J., Dela Paz, I., Riss, C., & Panahon, C. (2019). Applying spaced practice in the schools to teach math vocabulary. *Psychology in the Schools, 56*(6), 977-991.

Pyc, M. and Rawson, K., (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition, 35*(8), 1917-1927.

Pyc, M., & Rawson, K. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*(4), 437–447.

Rau, M. A., Aleven, V., & Rummel, N. (2010). Blocked versus interleaved practice with multiple representations in an intelligent tutoring system for fractions. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent tutoring systems* (pp. 413–422). Berlin/Heidelberg, Germany: Springer-Verlag.

Rau, M. A., Aleven, V., & Rummel, N. (2013). How to Use Multiple Graphical Representations to Support Conceptual Learning? Research-Based Principles in the Fractions Tutor. *AIED*. 762-765.

Rawson, K., & Dunlosky, J. (2012). When is practice testing most effective for improving the durability and efficiency of student learning? *Educational Psychology Review*, *24*(3), 419–435

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning*, *4*, 11-18.

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20*, 1209–1224.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*, 481-498.

Rohrer, D., Dedrick, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review*, *21*, 1323- 1330.

Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology*, *107*, 900-908.

Rohrer, D., Dedrick, R. F., & Hartwig, M., (2020a). The Scarcity of Interleaved Practice in Mathematics Textbooks. *Educational Psychology Review*, *32,* 873-883.

Rohrer, D., Dedrick, R. F., Hartwig, M. K., & Cheung, C.-N. (2020b). A randomized controlled trial of interleaved mathematics practice. *Journal of Educational Psychology, 112*(1), 40–52.

Schutte, G., Duhon, G., Solomon, B., Poncy, B., Moore, K., & Story, B. (2015). A comparative analysis of massed vs. distributed practice on basic math fact fluency growth rates. *Journal of School Psychology, 53*(2), 149-159.

Smith, S. M., & Rothkopf, E. Z. (1984) Contextual Enrichment and Distribution of Practice in the Classroom, *Cognition and Instruction*, *1*(3), 341-358.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24*(6), 837-848.

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation*. *60*(1). (pp. 113–189). Waltham: Academic Press

Yazdani, M. A., & Zebrowski, E. (2006). Spaced reinforcement: an effective approach to enhance the achievement in plane geometry. *Journal of Mathematical Sciences and Mathematics Education*, *1*, 37-43.

Ziegler, E. & Stern, E., (2014). Delayed benefits of learning elementary algebraic transformations through contrasted comparisons. *Learning and Instruction*, *33*, 131-146.

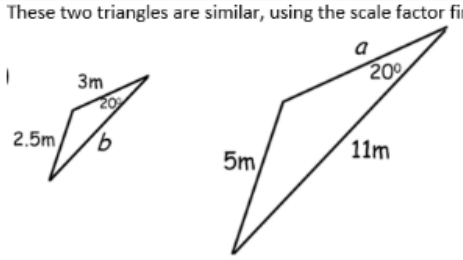
Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*, 750–763.

7. Appendices

Appendix A

A copy of one of the interleaved practice worksheets used in the lesson.

These two triangles are similar, using the scale factor find a and b



Complete the table of values for y=2x+3

X	-2	-1	0	1	2
y	-1		3		

Give the coordinates you would plot:

Solve the following simultaneous equations:

$$2x - 5y = 11$$

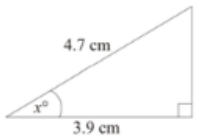
$$x - 4y = 4$$

Work out the value of $(3 \times 10^7) \times (9 \times 10^6)$
Give your answer in standard form.

Expand (x+3)(x-5)

Factorise

$x^2 + 10x + 16$



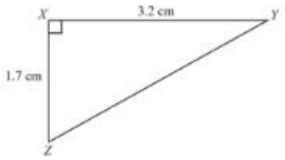
Work out the value of x.
Give your answer correct to 1 decimal place.

Make x the subject

$\frac{x+3}{z} = a$

Use ruler and compasses to **construct** the perpendicular bisector of the line AB.

You must show all your construction lines.



XYZ is a right-angled triangle.
$XY = 3.2$ cm.
$XZ = 1.7$ cm.

Calculate the length of YZ.

Give your answer correct to 3 significant figures.

Here are the first five terms of an arithmetic sequence.

−1 3 7 11 15

(a) Find, in terms of n, an expression for the nth term of this sequence.

.....

In another arithmetic sequence the nth term is $8n - 16$

John says that there is a number that is in both sequences.

(b) Explain why John is wrong.

.....

Joe travelled 60 miles in 1 hour 30 minutes.

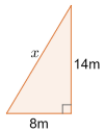
Work out Joe's average speed.
Give your answer in miles per hour.

Appendix B

The test given to both classes after the intervention. Both classes set the test in a 50-minute lesson with the papers being cross marked by myself and the other teacher:

Question 1

Work out the value of x. Give your answer to 3.s.f.



[3 Marks]

Question 2

It take 5 people 6 hours to renovate a garden. How many people would you need for the renovation to take only 2 hours?

[2 Marks]

Question 3

Rearrange the following formula to make x the subject

$4 + y = 3x - 5y$

[2 Marks]

Question 4

By working out the nth term rule, give the 178th number in this sequence.

15, 9, 3, -3, -9....

[3 Marks]

Question 5

The distance from the Earth to the moon is 3.844×10^6 m. A space craft travels at 9.61×10^{-7} m/s. How long does it take for the space craft to travel to the moon. Give your answer in standard form, and give a suitable unit for your answer.

[3 Marks]

Question 6

Expand

$(a + 4)(a - 7)$

[2 Marks]

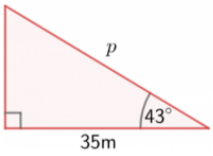
Question 7

Using ruler and compass, in the space below, accurately construct a 90 degree angle.

[2 Marks]

Question 8

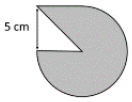
For the following triangle show that the value of p is 47.9m (3s.f.)



[3 Marks]

Question 9

Find the area of the following shape, giving your answer to 3 decimal places



[4 Marks]

Question 10

Solve the following equations

A) $\frac{3x}{4} + 10 = 2$

[2 Marks]

B) $\frac{2(3x+4)}{3} = \frac{24x-8}{2}$

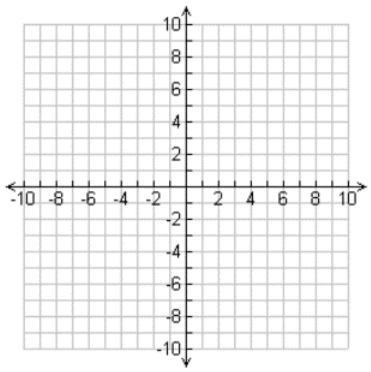
[4 Marks]

Question 11

a) On the coordinate axes below, plot the line

$$y = 2x - 6$$

[3 Marks]



b) Alfonso says that the line $6x + 3y = 24$ is parallel to $y = 2x - 6$. Explain why he is wrong.

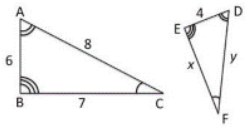
[2 Mark]

c) Give the equation of a line that goes though the points (2, -8) and (-4, -5)

[4 Mark]

Question 12

Given that the following two triangles are similar, find the values of x and y



[3 Marks]

Question 13

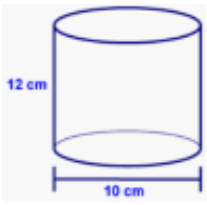
Solve the following simultaneous equation for x and y. Solutions with no working will be given 0 marks.

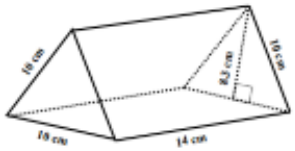
$$2x + 10y = 6$$
$$3x - 5y = 25$$

[3 Marks]

Question 14

Geoffrey is going to make a hollow door stop out of sheets of metal. The measurements he can construct his door stop are as follows





How much more material will Geoffrey need to make the cylindrical version compare to the triangular prism?

Appendix C

The raw data from the question by question analysis made of the test results.

1 in the table indicates that the pupil answered that questions correctly in the assessment, thereby gaining full marks for the question. 0 in the table means the pupil did not gain all marks, they scored anything less than full marks.

n is the percentage of pupils who correctly answered the question.

		Question Topic													
		Pythagoras	Proportion	Rearranging	Sequences	Standard Form	Expanding	Constructions	Trigonometry	Area	Solving	y=mx+c	Similar shaoes	simultaneous	surface area
Class A	A	0	0	0	0	1	1	0	1	0	0	0	1	1	0
	B	1	1	1	0	0	1	0	0	0	1	0	0	0	0
	C	0	0	1	1	0	0	0	0	0	0	0	0	1	0
	D	0	0	0	0	0	0	0	0	1	0	0	1	0	0
	E	0	0	0	0	1	1	0	1	0	0	0	0	1	0
	F	1	1	1	0	1	0	1	1	0	0	0	1	1	0
	G	1	1	0	0	0	1	0	0	0	0	0	0	1	0
	H	1	0	0	0	0	0	1	1	0	0	0	0	0	0
	I	1	0	1	1	1	1	0	0	1	0	0	0	0	0
	J	1	1	1	0	1	1	1	0	0	1	1	1	1	0
	K	1	1	1	1	1	1	0	1	0	0	0	0	1	0
	L	0	1	0	0	0	1	0	0	1	0	0	0	0	0
	M	0	0	1	1	0	0	1	1	1	0	1	1	1	1
	N	1	0	1	0	0	0	1	0	0	1	0	0	0	0
	O	1	1	0	0	1	0	0	0	0	0	0	0	1	0
n		60%	47%	53%	27%	47%	53%	33%	40%	27%	20%	13%	33%	60%	7%
Class B	A	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	B	0	1	0	0	0	0	1	0	0	1	0	0	0	0
	C	0	0	1	0	0	0	0	0	0	0	0	0	1	0
	D	1	0	1	1	0	0	1	0	0	1	0	0	0	0
	E	1	1	0	0	1	0	0	0	0	0	0	1	0	0
	F	1	0	1	1	0	0	1	0	1	0	0	0	0	0
	G	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	H	1	0	1	0	1	1	0	0	1	0	0	0	0	0
	I	0	0	1	0	0	1	0	0	0	0	0	0	0	0
	J	0	0	1	0	1	1	0	0	0	0	0	0	0	0
	K	1	0	1	0	0	0	0	1	0	0	0	0	1	0
	L	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	M	1	1	1	1	1	1	1	0	1	0	0	1	0	0
n		54%	23%	69%	23%	31%	38%	31%	8%	23%	15%	0%	15%	15%	0%
Differnece (na- nb)		6.2%	23.6%	-15.9%	3.6%	15.9%	14.9%	2.6%	32.3%	3.6%	4.6%	13.3%	17.9%	44.6%	6.7%

Appendix D

An extract from the process of coding the transcript of the interview, highlighting the coding assigned to the comments made by the pupils in the interview.

Revision	Confidence	Understanding
Difficulty	Structure	
How are you finding it?		
A	B	C
I think it's really good to revise the topics that we have already done	Yeah I kind of agree you do somehting like simultaneous equations and then perpendicular bisector and theyre just like really different	I find it good to revise the topics that we get like stuff we go over
but with simltaenous equations I didn't get it so probably like doing it for the entire lesson for 50 minutes would probably be more helpful, just do it do it do it. Whereas doing 4 questions might not have helped	Review lessons were really handy because you can see how much you have forgotten and go over it, but you havent just like stopped what you're learning to do it.	but I find it confusing because we've just finished what we've learned and then we move on to that and your brain kind of changes completely on what you're working on. amd then when next lesson comes I kind of forget what we were doing before.
and then doing other things I don't know makes my brain hurt a bit	So I think we should do definitely more questions of the thing we're actually learning	
yeah I think the amoutn of uestions in the interleaved and simultaneous equaion is fine, but I think trying to do simultaneous equations the next day is hard.	If I'm finding it hard I want to be able to do more	
	I want the choice of moving on if I feel confident but staying if I don't	