DOCTOR OF PHILOSOPHY

Using evidence-based methods to support children 'at risk' of poor academic outcomes to develop their mathematics skills

Owen, Kaydee

Award date:
2020

Awarding institution:
Bangor University

Prifysgol Bangor

Bangor University

Using evidence-based methods to support children 'at risk' of poor academic outcomes to develop their mathematics skills

Kaydee Leanne Owen

Thesis submitted to the School of Psychology, Bangor University, in partial fulfilment for the degree of Doctor of Philosophy

September 2020

Declaration

Yr wyf drwy hyn yn datgan mai canlyniad fy ymchwil fy hun yw'r thesis hwn, ac eithrio lle nodir yn wahanol. Caiff ffynonellau eraill eu cydnabod gan droednodiadau yn rhoi cyfeiriadau eglur. Nid yw sylwedd y gwaith hwn wedi cael ei dderbyn o'r blaen ar gyfer unrhyw radd, ac nid yw'n cael ei gyflwyno ar yr un pryd mewn ymgeisiaeth am unrhyw radd oni bai ei fod, fel y cytunwyd gan y Brifysgol, am gymwysterau deuol cymeradwy.

I hereby declare that this thesis is the results of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards.

Funding

The Economic Research and Social Council Doctoral Training Partnership (ESRC-DTP) funded the studentship that led to the completion of this research. Grant number: ES/P00069X/1.

Acknowledgements

This PhD would not have been possible without the support and input of many people, but there are a few who deserve a special thank-you. First and foremost, I would like to thank my project supervisors Professor Carl Hughes, Dr Richard Watkins, and Dr Michael Beverley. Carl, thank-you for providing me with lots of wonderful opportunities. Without you pushing me outside of my comfort zone over the last 4-years, I wouldn't have half the skills or confidence I have today. Richard, thank-you for helping me keep the SAFMEDS torch alight and for providing plenty of opportunities for me to disseminate my research. Mike, I will forever be indebted to you for helping me discover my passion for all things EBMEd. I would never have ended up on this path without your module and constant encouragement—thank-you for igniting my heart to chart! I also owe thanks to Professor Richard Hastings for his invaluable input, expertise, and feedback during the RCT.

Needless to say, I extend huge gratitude to the children, teachers, staff at XLP, and blind assessors who contributed to the data within each chapter—without you this thesis would cease to exist. Over the course of this PhD, I have been touched by the enthusiasm of teachers and children who have taken these methods on board. I will always be grateful for the experiences that led me to the researcher that I am today.

GwE have provided me with endless support with recruitment, training, and dissemination throughout the course of this PhD. To those I have worked with, it has been a real pleasure collaborating with you over the last few years and I'm excited to see where these methods go next.

Emma and Sarah, we have been on this journey together and I am so proud of both of you—thanks for always being my cheerleaders and spurring me on to get past the finish line. Stacey and Becks, thank-you for always being there for me when the journey ebbed. I have no doubt my combat gear and escape room knowledge will come in handy at some point in in the next chapter of life. Sion and Cameron, you have helped keep me sane and smiling even on days when it felt like there wasn't much to smile about. I don't think I can ever put into words how much I appreciate your kindness and friendship. Josh, thank-you for always being on hand when I needed a listening ear, a gentle nudge, and some stats advice. You are one of the most patient people I know, and I admire you so much for that.

My final thank-you goes to my family who have supported me through 7-years of university. Mum and dad, thank-you for being there every step of the way, reminding me what is important, and never letting me give up. Theo, you've only been on this earth for 6 years, but I am so unbelievably proud to call you my little brother. I hope finishing this PhD goes some way towards making you proud in return.

Contents

List of Figures

List of Tables

Overview of thesis

Chapter 1 outlines the existing literature and key terminology surrounding evidence-based methods within education; including Precision Teaching (PT), the Say-All-Fast-Minute-Every-Day-Shuffled (SAFMEDS) strategy, and Direct Instruction (DI). **Chapter 2** provides some additional context and justification with regards to the research methods we employed throughout this thesis. Here we also detailed the aims for the proceeding empirical chapters.

Chapter 3 details a cluster-randomized controlled trial assessing whether ongoing support from a researcher is necessary to yield the best outcomes from a teacher led SAFMEDS mathematics intervention. Following teacher training, we randomly allocated 33 schools to receive three in-house implementation support visits and ongoing email contact with a researcher. The remaining 31 schools received no implementation support following the same training. The results indicated that low-intensity support from a researcher has a positive effect on children's ability to recall addition, subtraction, multiplication, and division facts fluently (Mathematics Fluency and Calculation (MFaCTs): Grades 1-2, $d = 0.23$, 95% CI: 0.06 to 0.39; MFaCTs: Grades 3-5, $d = 0.25$, 95% CI: 0.08 to 0.43).

Previous quantitative studies have built up an evidence-base in support of using the SAFMEDS strategy in schools to promote fluency development across the curriculum. Yet, limited research has assessed the social validity key stakeholders associate with using the SAFMEDS strategy in schools. Using qualitative methods, **Chapter 4** outlines teachers' and children's views on the SAFMEDS strategy. In *study 1*, we disseminated an online survey to teaching staff ($N = 55$) who had attended SAFMEDS training in North Wales. Their responses revealed five themes relating to the advantages and disadvantages of adopting a teacher led SAFMEDS program in schools: (1) factors that promote progress, (2) factors that limit progress, (3) impact of competition, (4) confidence, and (5) inherent advantages of the SAFMEDS strategy. In *study 2*, we present data from interviews with children ($N = 26$) who had previously engaged with a SAFMEDS mathematics intervention program for at least one academic year. These transcripts revealed five further themes relating to the social validity of the strategy: (1) enjoyment, (2) data, (3) sense of achievement, (4) skills, and (5) home use.

XLP are a London-based charity who initiated a literacy and numeracy program (XL-LAN) in 2016. This program aimed to support young people from some of the most deprived areas in London to develop their basic literacy and numeracy skills through the SAFMEDS strategy. XLP recruited the expertise of researchers at Bangor University to help train, support, and develop the program. **Chapter 5** outlines the process of this collaboration, the key results from the pilot scheme, and some recommendations for future practice. We analyzed the children's ($N = 263$) in-session SAFMEDS data over the duration of their involvement with the program. The results indicated that children make session-by-session progress when they use the strategy, and this progress appears to be more stable for literacy skills compared to numeracy. We also identified some differences in the children's performance progress at the level of supervision/provision. Following interviews with children who had been involved in program ($N = 38$), we identified five themes relating to the social validity of the program: (1) procedure, (2) improvement, (3) revision tool, (4) home use, and (5) withdrawal from class.

Chapter 6 outlines a feasibility study evaluating the effectiveness of an instructional fluency approach (combining DI and PT approaches) to teach addition skills in a pupil referral unit. This study focuses on the progress of five boys who had been excluded from mainstream education. Over six school weeks, a researcher worked through a commercially available DI program (*Corrective Mathematics*) and randomized fluency practice sheets with each child on a one-on-one basis (adhering to PT methods). The results suggested that this intervention approach can help remediate early mathematics skill deficits and appears to be of most benefit to children who attend and engage with sessions regularly.

Chapter 7 provides an overview of chapters 2 through 5, highlighting some of the strengths, limitations, and applications of this research. We also present some ideas for future research in this field.

Chapter 1 : Introduction

The importance of mathematics and numeracy skills

The term *mathematics* refers to an international discipline which integrates concepts, rules, and procedures involving quantities and symbols. This includes learning about the arithmetic operations (addition, subtraction, multiplication, and division) and mathematical functions (Resnick & Ford, 1981/2008). Reyna, Nelson, Han, and Diekmann (2009) defined *numeracy* as the ability to understand and use mathematical information. In order to be considered numerate, a person needs to be able to connect the mathematics skills that they have learnt in a classroom to real-world problems (Geiger, Goos, & Forgasz, 2015). The development of a basic mathematics and numeracy skills aids functioning within everyday life; for example, basic addition and coin recognition skills allow us to go into a shop and pay for essential items (Ojose, 2011; OECD, 2016a). From an academic perspective, many further education courses and employment opportunities state that applicants are required to have adequate numeracy skills and mathematics qualifications (House of Commons, 2014). The real-world applicability of mathematics and numeracy highlights the importance of helping children to develop a competent skillset as they progress through school.

"At risk" groups and the attainment gap

Perry (2016) claimed that we can attribute a significant proportion of variation in school performance to children's background and their home learning environment. Whilst there are several factors that may influence attainment, there are two groups that are of specific focus within this thesis: children who are eligible for free school meals and children who attend pupil referral units. Internationally, children from lower socio-economic backgrounds are at increased risk of poor academic outcomes across the curriculum (OECD, 2011). Eligibility for Free School Meals (eFSM) is a parameter used to assess levels of deprivation and social disadvantage in school age children across the United Kingdom (UK; Department for Education, 2016a). Nationally, children who are eFSM are more likely to underachieve in mathematics and numeracy compared to their peers who are not eFSM (Dowker, 2016; Gorard, 2012; Perry, 2016; Taylor, 2017). Parliament UK (2014) claimed that this attainment gap is evident at preschool age and remains consistent throughout a child's education. Within the context of Chapter 3, it is important to consider the effect that this parameter could have had on the children's test performance. As such, we took this variable into account when analyzing the wider data set by adjusting for it as a fixed effect within the linear mixed effects model.

Within Chapter 6, we aimed to explore the effects of a remedial intervention in a pupil referral unit (PRU). PRUs accommodate children with complex needs (relating to behavior or illness) that cannot be managed within mainstream school settings (Estyn, 2015). A high percentage of children accessing alterative provision placements in PRUs have additional learning needs (approximately 75%), with many of the children also experiencing social, emotional, and behavioral difficulties (Department for Children, Schools and Families, 2008). Placements within a PRU can be either on a singular (full-time) or dual (part-time) registry basis depending on the nature of the child's exclusion from mainstream education. Many children who attend PRUs do so on a dual registry basis—meaning that they spend part of their school week in a PRU and the remaining time in mainstream school (Department for Education, 2017). This raises some challenges with regards to the unique learning experiences and skill knowledge that children present with upon registry.

In the academic year 2014/15, only 17% of children attending PRUs in England left education with a A*-C grade GCSE in mathematics (Ofsted, 2016). In a recent report, Estyn (2017) revealed that in four in ten PRUs in Wales children did not develop the skills they needed to support their learning across the curriculum. They also highlighted that the curriculum lacked challenge, did not ensure that all children in the classroom achieved, and that the pace of learning was too slow. This provides some national context to identify effective pedagogical approaches that enable teachers to support mathematical learning in PRUs and remediate some of the skill deficits that the children may present with.

Heckman, Pinto, and Savelyev (2013) investigated the long-term outcomes associated with the use of early childhood intervention programs for children from at-risk groups. They collected data over 35-years to provide strong evidence to suggest that high-quality early intervention programs can considerably improve outcomes in education, employment, and health. As well as having clear social significance for individuals, these programs also have long-term benefits in relation to strengthening the economy. That is, by investing in early education provision children do better in school, go into employment, and are less likely to be involved in crime. Consequently, early intervention leads to fewer individuals claiming state benefits and leaves the criminal justice system with lower expenditure.

Mastery-based versus fluency-based learning

Within educational practice, mastery reflects children being able to perform skills to a level of at least 80% accuracy (Dalton & Hannafin, 1988; Fuller & Fienup, 2018). However, this criterion does not reflect the *rate* at which children are able to perform these skills. Binder, Haughton, and Bateman (2002) highlighted that if classroom tasks enabled children to practice skills beyond traditional mastery, they will be able to recall facts *fluently* (i.e., accurately and at speed).

The literature conceptualizing how children learn suggests that there are two core stages involved in being able to perform skills competently—acquisition and practice (Kubina & Wolfe, 2005; Miller & Heward, 1992). When learning a new skill, a child first needs to understand *how* to perform it. Miller and Heward explained that during this stage teachers should direct the instructional activity towards teaching the child how to perform the skill accurately. Teaching during this stage is often slower-paced and it is important that the child receives feedback relating to their performance after each attempt that they make. Delay in this feedback is undesirable as it provides an opportunity for the child to practice making mistakes and learn bad habits (Van Houten, 1984). Once a child is able to perform the skill accurately, it is important that they continue to practice it. During this phase, Miller and Heward argued that the teacher and child should shift their focus towards fluency-building activities and the feedback should pertain to the child's rate of performance. Since the child should be able to respond accurately to most, if not all, of the questions during this phase it is not necessary to provide verbal feedback following each response, as this would impede the child's fluency.

There are several advantages of using fluency-based instruction, including the associated learning outcomes. Once a child is able to perform a skill fluently it is more likely that they will be able to satisfy four learning outcomes—retention, endurance, stability, and application (RESA; Binder, 1996). In turn, a child should be able: to perform the skill at the same level after a period of not practicing it; endure the skill for an extended timing period; maintain high levels of fluent performance in the presence of distractions; and generalize the skill to novel tasks without additional instruction. Advocates for fluency-based learning believe that there should be predetermined numerical benchmarks (or *fluency aims*) that predict the occurrence of achieving these learning outcomes (Johnson & Street, 2004; Haughton, 1980). For example, the fluency aim for seeing an arithmetic fact and writing the answer is 80 to 100 correct answers per minute (Johnson & Street, 2012).

Fluency-based learning may also have wider implications within the classroom with regards to how children perceive mathematics. Cates and Rhymer (2003) investigated the relationship between mathematical fluency and mathematics anxiety. A sample of college students completed timed tests in basic arithmetic operations. Those who measured high on a mathematics anxiety scale demonstrated consistently lower levels of fluency across the timed tests for each of the four arithmetic operations compared to those with low mathematical anxiety, despite there being no difference in the error rates between these two groups. These findings suggest that mathematics anxiety may be related to the level of learning (fluency) rather than overall performance (accuracy).

Complementary research also demonstrates the advantages of employing fluency-based criteria in the classroom over mastery-based. For example, Jordan, Hanich, and Kaplan (2003) administered mathematics tests to children aged 7 to 9 years. These tests included questions assessing arithmetic operations, story problems, approximation, calculation principles (i.e., understanding the relationship between the arithmetic operations), place value, forced retrieval of number facts, and written computation. All of these skills, with the exception of place value, involved the operations of addition and subtraction. The group of children who were able to answer arithmetic facts fluently consistently performed more accurately across these tests compared to age-matched peers who were less fluent. This suggests that fluency in arithmetic skills may impact a child’s ability to perform complementary and higher-order skills within the curriculum.

Amongst the behavioral science literature, Johnson and Layng (1996) identified that children need to acquire fluency of basic skills before progressing onto higher-level skills within any academic subject. If a child has a deficit in terms of acquisition and fluency of lower order skills within a curriculum, they will find it increasingly difficult to learn higher order content. For example, if a student cannot complete equations containing basic integer addition and subtraction then they will find it more difficult to master more complex calculations, such as those containing fractions (O’Connell & SanGiovanni, 2011). Binder (1996) coined this process as *cumulative dysfluency* and maintained that it can explain academic underachievement and failure within education. Across the UK, teachers follow a prescribed curriculum (see, for example, Department for Education, 2013; Welsh Government, 2016). This means that children are expected to master skills based on their chronological age each academic year. The cumulative dysfluency theory has clear implications for children who cannot perform age-expected skills fluently by the end of the corresponding school year. This provides an avenue of focus with regards to identifying effective preventative and remedial interventions throughout the different stages of education. This may have particular implications for those populations who are statistically more at risk of underachieving.

In terms of educational practice in Wales, the incoming mathematics curriculum and complementary guidance also identifies the need to support mathematical fluency development in the classroom. The incoming Curriculum for Wales 2022 guidance identifies that children need to be able to fluently use the four basic arithmetic operations and understand the relationship between them (Welsh Government, 2019). The Education Endowment Foundation (2017) also recommend that schools support children to be able to recall arithmetic facts fluently. They acknowledged that without this knowledge children are likely to encounter difficulty understanding and using the mathematical concepts that appear later in the curriculum. This reform within education will allow children to practice skills beyond the acquisition phase of learning.

When considering how to implement time-trials to support fluency practice within the classroom, teachers and researchers may want to consider Miller and Heward's (1992) guidelines. These include: (1) keeping timings short; (2) conducting time trials every day; (3) making time trials fun (i.e., framing them as a game rather than a test); (4) encouraging children to beat their own scores; (5) maintaining a fast pace by telling children not to be afraid of making mistakes; (6) providing more questions than anyone could possibly answer within the time trial to prevent artificial ceilings on performance; (7) keeping a record of the children's progress; and (8) evaluating the effectiveness of the program by charting data and looking at it regularly. These guidelines have clear applications with regards to *Precision*

Teaching (PT) and *Say-All-Fast-Minute-Every-Day-Shuffled* (SAFMEDS) strategy, as outlined below.

Precision Teaching (PT)

Kubina and Yurich (2012) described PT as a system for defining, measuring, recording, and analyzing learning on a child-by-child basis. The PT approach provides a set of systematic procedures for measuring the fluency of behavior and facilitating data-driven decisions within the classroom (including the use of the *Standard Celeration Chart*; SCC). As such, PT serves to compliment and evaluate existing curricula (Binder & Watkins, 1990; White 1986).

Underlying principles of PT

Lindsley (1995) outlined the three founding principles that underlie PT: monitor fluency regularly, use-self-recording, and “the learner knows best”. Lindsley believed that it is not enough to report percentage correct when discussing learning. This is because fluency is a more sensitive measure and provides us with more insight into how well someone can perform a skill. The more frequently we measure fluency of performance, the more we can support our learners to reach their goals. Lindsley claimed teachers have found that changes in learning are much greater when the learners themselves take an active role in this process. As such, we should encourage learners to record their own data on an SCC and use it to guide their future study. The final principle, the learner knows best, refers to the idea that everybody’s learning experience is unique. Our learning experiences are shaped by our biology, interactions with the environment, and how we are taught specific skills. If a learner is progressing, the teaching is right for them. If their learning is not progressing, the instruction they are receiving or the task that they are doing needs to change.

The Standard Celeration Chart (SCC)

By providing teachers and learners with the necessary scientific knowledge and tools, Lindsley (1972) reasoned that they could discover what procedures and materials produced the greatest improvements in performance. The SCC (see Figure 1.1) is a tool that enables teachers and learners to assess just that. If a learner's progress is not desirable, then they can consider how to revise the components of instructional design. Since its conception, the SCC has become a hallmark feature of the PT approach and enables us to record, visualize, and interpret data using a standardized scale (Boyce, 2003; Calkin, 2005; Lindsley, 1995).

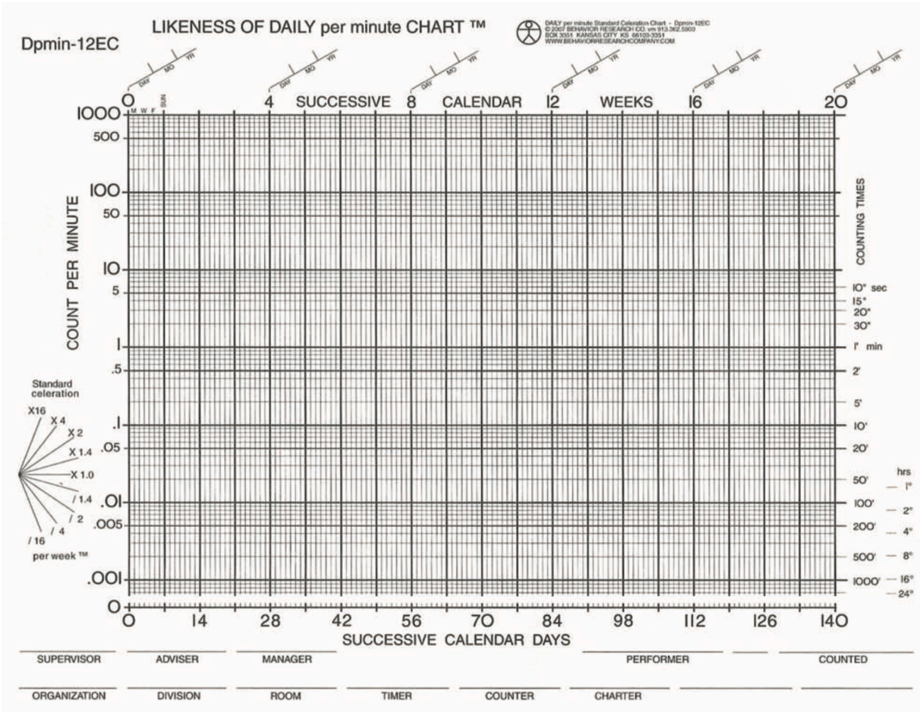

Figure 1.1. Image of an SCC as displayed in Street and Johnson (2014)

One of the defining features of the SCC is that the Y axis displays a ratio (multiply/divide) scale. Calkin (2005) explained the theoretical underpinnings of using a ratio scale to measure behavior over an equal interval (add/subtract) scale. In summary, it takes less time to achieve our desired goals if we work towards doubling the rate of desired behaviour per week, opposed to adding one each day. In terms of real-world application, multiply/divide scales more accurately approximate behaviour change and can help us to set realistic goals. Teachers and learners use the Y axis to plot the frequency of the behaviour they are recording. They can calculate this value by dividing the behaviour count (e.g., number of correct answers to arithmatic facts) by the counting period, in minutes (White & Neely, 2012). Over time we can see trends in the data displayed on the SCC, with regards to whether the frequency of behaviour is increasing (acclerating) or decreasing (decelerating; Lindsley, 1992).

The x-axis of the SCC is calendar-based, allowing learners to plot their performance over 20 consecutive weeks (Lindsley, 1991a). Each of the 140 days in this time period has its own vertical line on the chart, allowing for easy annotation. As well as plotting performance data, teachers and learners can evaluate the effects of breaks and school holidays on performance (White & Neely, 2005).

The standardization of the X and Y axis on the SCC makes it a universal tool for recording and visualizing data. Lindsley (1990) found that before the adoption of the SCC evaluating and communicating information about progress was a time-consuming process. This was largely because teachers would need to describe their unique recording and charting systems. Moreover, manipulation of scales and proportions can affect the visualization of progress, making non-standardized scales subject to bias (White & Neely, 2012). By using a standardized chart, such as the SCC, we can become accustomed to reading the same scales and interpret data with ease. This in turn increases the time that teachers can invest in evaluating their learners' progress, rather than explaining their charting practices (Lindsley, 1990).

An additional advantage of allowing teachers and learners to chart the frequency of classroom behavior against calendar days is that it enables users to implement interventions when they are necessary; preventing prolonged periods of non-progression (Hughes, Beverley, & Whitehead, 2007). Johnson and Street (2012) explained that the visual patterns of data on the SCC—that are formed by the celeration lines of the corrects and errors—reflect *learning pictures*. Lindsley (1995) outlined 13 learning pictures that can emerge from data on the SCC (see Figure 1.2). If a learner's correct responses are accelerating at the desired trajectory (i.e., displaying an improving picture) the instruction and materials are appropriate for their learning, and intervention is not required. If learners' data shows that they are making little or no progress towards their aim (i.e., their data is displaying a maintaining or worsening picture), then a change to instruction or materials may be necessary.

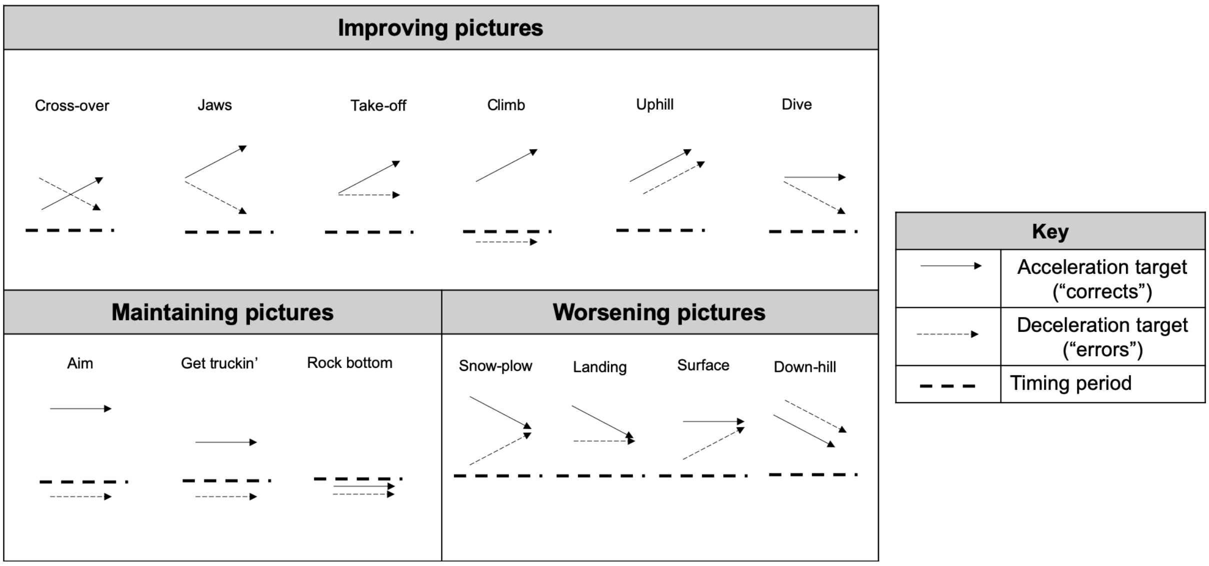

Figure 1.2. Lindsley's (1995) learning pictures based on the visualization of data on a SCC

Aninao, Acevedo, Newsome, and Newsome (2015) explained that by reducing the latency between collecting data and charting it, teachers and learners can make data-driven and real-time decisions about interventions. They maintained the assertion that the SCC is a tool that allows us to concurrently assess learning and decide when an intervention is necessary. If a learning picture suggests that a learner is maintaining a skill or their rate of responding is decreasing, for three consecuative sessions, then we should consider making a change to support the learner's progression.

Stages of a PT program

PT programs follow five strategic steps to help ensure learners are developing their skills: pinpoint, prepare, collect data, plot data, and make data-based decisions (Johnson & Street, 2004). During the pinpointing step, the teacher should identify what skill they want their learner to develop, how they are going to measure it, and what level they want their learners to be able to perform this skill at (i.e., they should set a fluency aim; for more information about some of the predetermined fluency benchmarks for mathematics skills see Johnson & Street, 2012). Once the skill has been defined, the teacher needs to identify and prepare the resources that their learner will need to assess it. The learner should then practice the pinpointed skill within a specified time sprint. Typically, timings vary between 1- and 5-minutes in length. The learners will then take their best score for the day and plot it on their SCC. Once a learning picture begins to emerge, the teacher and learner can review the data and make decisions about how to proceed (i.e., keep following current practice or implement an intervention).

Review of existing PT research

Although sparse, the existing empirical literature supports the idea that PT approaches can be integrated into classrooms within mainstream primary schools to support children's mathematical fluency development. Researchers have supported teachers to integrate the approach into their classrooms to support small groups of children (see Chiesa & Robertson, 2000; Gallagher, 2006; Strømgren, Berg-Mortensen, & Tangen, 2014). In these studies, the differences between pre- and post-test data provide evidence to suggest that applying PT to paper-based tasks can help children notably to improve their mathematics skills within a short timeframe (8-12 weeks). This literature reflects the benefits of using PT to support typically developing children to write digits (Chiesa & Robertson, 2000; Gallagher, 2006) and answer arithmetic facts fluently (Chiesa & Robertson, 2000; Strømgren et al., 2014).

Several empirical studies have also reported the benefits of using PT to support the mathematics learning of primary-aged children with additional learning needs. This includes children with unspecified learning disabilities (Peterson, Hudson, Mercer, & McLeod, 1990) and children who hold a diagnosis for attention deficit hyperactivity disorder (ADHD; Brady & Kubina, 2010). The results from these studies demonstrate improvements in fluency of the children's written recall of multiplication facts (Brady & Kubina, 2010) and verbal recall of place value (Peterson et al., 1990). Many of the studies that have implemented PT strategies to support mathematical fluency development amongst samples of children with additional learning needs represent samples of less than 10 children (for a review please refer to Ramey, Lydon, Healy, McCoy, Holloway, & Mulhern, 2016). Whilst this research is still useful in providing us insight into the benefits of PT with these populations, further replications may be necessary to demonstrate the reliability of this approach.

It is worth noting that these published papers may not be the only data available to promote the use of PT in the classroom to support fluency of mathematics skills. Proponents of PT have suggested that practitioners have shared SCC data informally at events rather than through formal peer-reviewed publications. Binder (1996) explained that from the 1970s, laboratories have held open chart-sharing sessions for people to informally report their data and discoveries to others. SCC sharing sessions have since appeared in convention programs and applied behavior analysis (ABA) conferences; providing regular open events for people in the field to discuss fluency-based projects (Lindsley, 1991b). Calkin (2002) estimated that as of February 2000, practitioners have used over 1-million SCCs to record data. This estimation reflects self-reported usage of the SCC from projects that practitioners have supervised or completed independently. To the author's knowledge this is the last recorded statistic on the usage of SCCs in this context, but still reflects the idea that practitioners use the SCCs more than publication records reflect (i.e., beyond research efforts).

Say-All-Fast-Minute-Every-Day-Shuffled (SAFMEDS)

Procedure and theory

Learners can apply the principles of PT to the SAFMEDS strategy. The SAFMEDS strategy serves as a learning tool that helps teachers to provide their learners with the materials they need to practice and assess basic fact-based skills. It also enables learners to fluently recall these facts at a performance level that promotes retention, endurance, stability, and application (Graf & Auman, 2005; Johnson & Layng, 1992). Learners engage with the SAFMEDS strategy using a deck of flashcards with a stimulus on the front (such as a question or a statement with a missing word), and the corresponding correct response on the back (Graf & Auman, 2005; Meindl, Ivy, Miller, Neef, & Williamson, 2013). Please refer to Figure 1.3 to see an example of a SAFMEDS card.

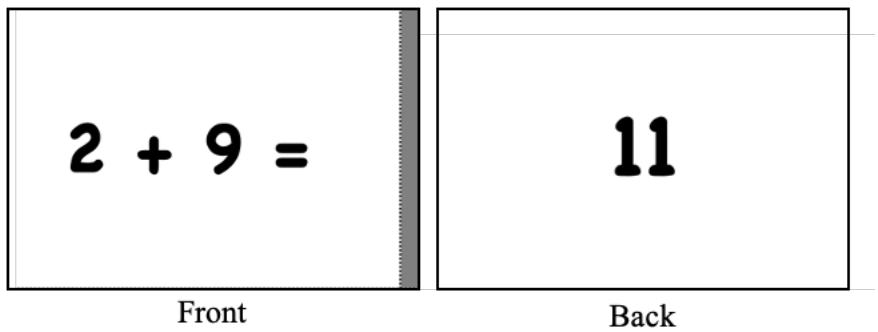

Figure 1.3. Example of a SAFMEDS card

Lindsley (1992) and Graf and Auman (2005) explained that each letter of the SAFMEDS acronym aims to prevent learners from making the most common errors in fluency building whilst using the cards. Learners should: (1) *say* the answers to prevent silent viewing, making it easier to validate responses; (2) use *all* the cards in the deck, rather than learning them in sections and collating them later; (3) work through the deck *fast* so they do not need to focus on building fluency later; (4) see how many cards they can get through during one *minute* sprints to keep timings brief and focused; (5) run timings *each day* to create habits and eliminate the need for cramming before tests; and finally (6) *shuffle* the cards between timings to prevent serial learning. Graf and Auman also provided additional guidance to support teachers and their learners during SAFMEDS sessions. For example, learners should read the front of the card silently to aid fluency and then verbalize the answer. Only after they have answered out loud should they turn the card over to check their response; this prevents them simply reading the answer and promotes anticipation. Receiving immediate feedback from the back of the card is an important aspect of the SAFMEDS approach. Lindsley (1996a) explained that by reducing the delay between the learners' response and feedback, it helps to prevent continual rehearsal of the incorrect response.

During each timing, learners should separate their cards into two piles; corrects and errors (Cihon, Strutz, & Eshleman, 2012). McGreevy (1983) maintained the idea that learners should not be punished for their mistakes, as they are simply "opportunities to learn". As such, many proponents of PT approaches refer to errors or hesitations as *learning opportunities* (e.g., Beverley, Hughes, & Hastings, 2009; Cihon et al., 2012; Hughes et al., 2007; Merbitz, Vieitez, Hansen Merbitz, & Binder, 2004). After each SAFMEDS timing, learners can go through their learning opportunities pile and ask their teacher questions about the content of the cards that they do not understand. Graf and Auman (2005) explained that teachers should teach to their learners' requests rather than spending time explaining content that is self-explanatory and/or their learner has already mastered.

The SAFMEDS strategy is designed in such a way that encourages learners to complete many of the aspects independently. Lindsley (1996b) detailed the importance of freeing the operant during fluency trials to keep the rate of responses high. That is, we should ensure that our tasks and materials do not place any ceilings on our learners' performance, so that they can answer questions at their own pace. During SAFMEDS timings, learners should work through a deck with more cards than they can possibly get through within one minute to avoid any measurement-defined or procedure-imposed ceilings (Binder, 2010; Johnson & Street, 2004). Graf and Auman (2005) provided advice surrounding how to create flashcards that were suitable for SAFMEDS timings. This advice includes avoiding wordy stimuli on each card and providing a discriminative marking (such as a grey line) on one side of the cards to allow learners check their cards are facing the right way around before each timing. To further promote autonomy and free the operant, learners should hold and direct their own cards. Lindsley (1996b) found that learners are able to establish a rhythm which leads them to present the next card to themselves twice as fast as a partner-directed approach could.

Review of existing SAFMEDS research

Quigley, Peterson, Frieder, and Peck's (2018) systematic review revealed that 27 peer-reviewed data-based articles reported the effects of the SAFMEDS strategy. Of these articles, only four reported learners using the strategy to support arithmetic skills. These studies adopted small *N* within subject designs with children with a learning disability attending mainstream schools (Casey, McLaughlin, Weber, & Everson, 2003; Cunningham, McLaughlin, & Weber, 2012), children with neuropsychiatric conditions attending a residential setting (Hartnedy, Mozzoni, & Fahoum, 2005), and children who have sustained a traumatic brain injury (Chapman, Ewing, & Mozzoni, 2005). Due to the nature of these efficacy studies, a researcher delivered and maintained the programs and in some cases the learners received supplemental support, such as pre-timing practice (Casey et al., 2003; Cunningham et al., 2012; Hartnedy et al., 2005) and instruction to correct errors between each timing (Hartnedy et al., 2005). All of these studies demonstrated that the SAFMEDS strategy can elicit positive fluency gains in arithmetic skills.

Since the Quigley et al (2018) review, several further studies have emerged. These studies shifted towards group design paradigms and provided further insight into the scalability of SAFMEDS programs in mainstream schools to support fluency of mathematics skills. Hunter, Beverley, Parkinson, and Hughes (2016) investigated the effects of class-wide implementation of the SAFMEDS strategy compared to teaching as usual. Their results demonstrated that all of the children who completed SAFMEDS timings at the end of their mathematics lessons could answer questions about percentages of money significantly more fluently after using the cards for four-weeks. This group also maintained this fluency after a month of no explicit practice. The children who did not engage with the SAFMEDS strategy also made sustained improvements, but the pre- to post-test differences were not as significant as the SAFMEDS group. Greene, Mc Tiernan, and Holloway (2018) provided an additional example of the practicalities and applicability of using the SAFMEDS strategy in mainstream classrooms. In their study, Greene et al. explored whether older children (10-12 years old) could tutor younger peers (8-10 years old) through their SAFMEDS timings and the error correction procedure. The experimental group participated as tutees in the SAFMEDS program for 30 minutes, three-days per week, outside of their scheduled mathematics lesson. Analysis revealed that the cross-age peer tutoring approach helped children to significantly improve their fluency of arithmetic facts over just eight-weeks compared to the results of those who received no additional mathematics tuition.

Both Hunter et al. (2016) and Greene et al. (2018) provided promising results that support the use of SAFMEDS in mainstream educational settings. However, it is worth noting that in both studies a researcher was present at each SAFMEDS session to ensure procedural fidelity and support data-driven decisions. To the author's knowledge Beverley, Hughes, and Hastings' (2016) study is the only published paper that reports teacher led implementation of a SAFMEDS program; whereby a member of the research team was only present to administer the pre- and post-tests but were not in situ during the program itself. The results from their study suggest teachers are able to support a class of children to implement the SAFMEDS strategy to elicit positive fluency gains in second language vocabulary. In Chapter 3, we explored teacher led implementation further to investigate whether coaching from a researcher with experience with the SAFMEDS strategy is necessary to yield the greatest fluency gains in the classroom. Moreover, in Chapter 5 we report the results of an exploratory impact study whereby a youth-based charity coached teachers and mentors to implement a SAFMEDS program in range of educational settings.

The vast majority of published research reporting the effects of the SAFMEDS strategy focus on the quantitative fluency development of learners. Beyond these outcomes, Fawcett (1991) highlighted the importance investigating the social acceptability of an intervention's goals, procedures, and effects. As part of the research process, researchers should investigate whether key stakeholders will accept and use intervention. Moreover, researchers should assess whether the intervention is still viable when stakeholders (with less expertise in research and implementation) use it within an applied quasi setting (Schwatz & Baer, 1991). Some anecdotal reports claim that children enjoy using SAFMEDS as a method of learning (Beverley et al., 2016) and would choose to continue using it after the termination of a research study (Hunter et al., 2016). In Chapter 3, we present two qualitative studies that aimed to gain further insight into teachers' and children's experiences of using the SAFMEDS strategy in their schools.

Direct Instruction (DI)

Theory

Similar to PT, proponents of DI believe that all children can learn if their teachers provide them with appropriate instruction and materials (Stockard, Wood, Coughlin, & Khoury, 2018; Flores & Kaylor, 2007). The initiation of DI was a systematic attempt to build a technology that would help children learn new skills in the most time-efficient way possible (Engelmann, Becker, Carnine, & Gersten, 1988). DI programs build upon the assumption that children can learn new material when they have mastered prerequisite skills to fluency and the instruction is unambiguous (Stockard et al., 2018). Slocum (2004) used the analogy of a staircase to explain instructional programs. He explained that to be effective and inclusive, DI programs should enable as many children who arrive at the bottom of the staircase (i.e., those who have the prerequisite skills) to reach the top (i.e., to master the objectives) as possible. To get from the bottom to the top of the staircase, children must complete each step. There are five main strategies used within DI programs that make progress up the staircase as simple as possible: (1) clear and explicit instruction, (2) instruction sequencing, (3) sufficient practice and mastery criteria, (4) intention to fade support, and (5) clear instructions for teachers.

Clear and explicit instruction. DI lessons focus on teaching children cognitive learning; this includes teaching rules, concepts (i.e., skills taught by providing learners with a definition), and strategies (i.e., skills that require learners to enact steps in a specific sequence; Kozioff, LaNunziata, Cowardin, & Bessellieu, 2000). To teach concepts and rules effectively, teachers much explain them to their learners clearly and directly. Kinder and Carnine (1991) explained that teachers should illustrate what they have verbally stated by using an array of true examples and non-examples. True examples are stimuli that vary on irrelevant attributes, but they should maintain the same critical features. On the contrary, non-examples are stimuli that are similar to one and one another but each miss a critical feature. Teachers should present examples and matched non-examples one after the other. By doing this, they force the learner to attend to the small changes between the two stimuli and identify the critical features, as all other features remain consistent.

One of the driving principles underlying DI is that teachers should organize the curriculum in such a way that teaches generalizable strategies (Engelmann & Carnine, 1982). That is, they should teach the skills, concepts, and other knowledge structures that allow learners to go beyond the examples taught and apply their knowledge to novel situations.

These foundations help educators to organize the content of DI curricula and design the scripts that teach this content (Kame'enui, Carnine, Dixon, Simmons, & Coyne, 2002).

Instruction sequencing. Before introducing new concepts, rules, or strategies, teachers much first ensure that their learners have mastered any associated skills (Slocum, 2004). To do this, Kozioff et al. (2000) explained that the curriculum developer must identify the basic (e.g., number recognition, place value, single digit addition) and complex skills (e.g., addition sums involving carryovers, algebra) that contribute to the overall knowledge system (e.g., mathematics). The curriculum developer then needs to design a lesson with precise wording and examples for each skill that a learner needs to master. Each lesson aims to help learners acquire new knowledge, review key ideas, and synthesize their knowledge across the curriculum

Sufficient practice and mastery criteria. Most DI programs incorporate placement tests to help teachers to determine the skills that a learner has already mastered. By administering these tests, teachers are able to ensure that they deliver instruction at an appropriate skill level. The lessons should not include material that is either too challenging or repetitive of the skills that a child has already mastered (Stockard et al., 2018; Moran & Malott, 2004). If teaching at a group level, these assessments help teachers to place their learners into small instructional groups based on their skill level. By tailoring instruction to the strengths and needs of the group, all learners have the best chance of learning all of the material. Teachers are able to re-evaluate these groupings regularly throughout the program (through both the lesson activities and mastery tests). They can alter children's membership to a group if they are mastering the content faster than their peers (Kozioff et al., 2000).

Small group instruction can be advantageous as it is more time efficient than one-on-one teaching and places an added emphasis on oral communication (Kinder & Carnine, 1991). Each child should have multiple opportunities within each lesson to engage in verbal exchanges with their teacher, to allow for increased engagement and continual progress monitoring. In group settings, this is achieved by children responding in unison when their teacher provides a signal (Wolery, Ault, Doyle, Gast, & Griffen, 1992). This means that *all* of the children in the group have lots of opportunities to respond throughout, rather than relying on the same children in a class to answer questions (Moran & Malott, 2004).

The ultimate aim of an instructional program is to help learners to master the skills within the curriculum. Like in PT, DI programs help children to be able to perform skills independently and fluently (Kozioff et al., 2000; Kamps et al., 2004). Englemann (2007) stated that in order for program design to support mastery, lessons should not present lots of new information. New concepts, rules, strategies, and training should only account for 10-15 percent of the total lesson. The rest of the lesson should focus on firming up and reviewing the material that the children have learnt earlier in the program. This design provides children with plenty of opportunities to practice the skills that they have learnt in the preceding few lessons, to consolidate and apply new skills, and build fluency.

Intention to fade support. In line with Lindsley's (1996a) and Van Houten's (1984) beliefs, proponents of DI believe that teachers should minimize the delay between a child's response and feedback; particularly if a child makes a mistake. To help minimize and remedy errors, DI developers generally follow the model-lead-test-delayed test method (Kozioff et al., 2000). Bechtolt, McLaughlin, Derby, and Blecher (2014) explained that first a teacher should model the skill they want the child to learn (e.g., state the arithmetic fact "1 add 2 equals 3" orally). The teacher and child should then perform the skill together, in unison. To assess that the child has understood, they should then perform the skill without their teacher providing leading questions and guided prompts. If the child makes an error, the teacher can repeat this process with them to build confidence and firm up the skill. As a teacher progresses through the model-lead-test-delayed test approach, they gradually reduce the intensity of the prompts they provide (Swanson, 1999). Over time this causes children to apply the strategies silently and independently (Archer & Hughes, 2011; Kinder & Carnine, 1991). The delayed testing element of this method involves children demonstrating retention and their understanding of the skill in subsequent lessons (Kozioff et al., 2000; Slocum, 2004).

Clear instruction to teachers. DI programs prescribe specific teaching procedures to ensure that teachers have the necessary materials to deliver content effectively and unambiguously. Kozioff et al. (2000) explained that a script accompanies each lesson so that teachers present exercises in a systematic order; with skills increasing in difficulty as the curriculum progresses. A lesson might begin with a teacher-directed segment (verbal). The teacher may then set the children a group or independent activity so that they have the opportunity to practice and generalize skills to new problems.

Teachers should deliver the contents of each lesson at a brisk pace (Kinder & Carnine, 1991). Carnine (1976) explained that by engaging in rapid question-answer exchanges, teachers are able to cover as much content as possible in each 30-45-minute lesson. The rate of exchange also helps to keep children engaged throughout. This is particularly important when considering children who are not performing at a level that reflects their chronological age and school year group. By engaging in DI sessions, these children should be able to make significant progress in the subject and work towards closing the attainment gap (Engelmann, 1999). The scripts also emphasize the importance of positive reinforcement throughout the instructional process (e.g., providing verbal praise for good listening, responding on signal, and providing correct answers). This makes the learning process rewarding for children and, by proxy, more rewarding for the teachers as they watch their group progress (Engelmann et al., 1988; Engelmann, 2014).

Importantly, teachers test scripts with small groups to validate that the instruction is unambiguous before being field tested at scale in schools. During the field-testing stage teachers provide detailed feedback regarding the problems that their students have faced whilst engaging with the program. This feedback feeds into the revisions before the designers release it for further field testing. Only when teachers and students report no further problems with the program it can be sent for publication, commercialization, and wider dissemination (Huitt, Huitt, Monetti, & Hummel, 2009; Kozioff et al., 2000).

Review of existing DI research

In the 1960s the federal government in the United States legislated a large-scale social project entitled '*Project Follow Through*'. This project aimed to evaluate the outcomes of curricular programs on three dimensions: basic academic skills (reading, arithmetic, spelling, and language), problem-solving skills, and self-concept (including developing positive attitudes towards learning). Specifically, this project evaluated programs that claimed to eliminate the discrepancies in achievement between children from disadvantaged backgrounds and their peers (Watkins, 1997). At its peak, Project Follow Through had 60,000 children enrolled onto one of the 20 programs on offer, with the sample representing varying demographics across 178 communities (Egbert, 1981). The final analysis reported the outcomes across nine curricular programs: DI, Parent education, Behavior Analysis (token economy, programmed instructional materials, data-driven decisions), Southwest Labs, Bank Street, Responsive Education, Tucson Early Educational Model (TEEM), Cognitive curriculum, and Open Education (Watkins, 1997). The results from this project favored DI across all three dimensions (Adams & Engelmann, 1996). However, after the publication of the results from Project Follow Through, House, Glass, McLean, and Walker (1978) highlighted several critiques. This included problems with the project's vague and ill-defined classification systems, biased assessments, and lack of sophistication with regards to the data analysis technique. Despite project follow-through being one of the largest education experiments to date, further research was necessary to establish the true impact of DI programs.

Stockard et al. (2018) conducted a comprehensive meta-analysis on studies and reports using DI methods between the 1960s and 2017. They defined a report as an individual publication (e.g., an article or dissertation) and a study as a data-gathering effort (e.g., an experiment) upon which a report is based. Their search identified 328 eligible studies and 393 eligible reports that focused exclusively on the delivery of a DI program in a school-based setting. By using mixed effects models, Stockard et al. were able to consider the impact of several variables on the reported effect sizes amongst the literature. As such they were able to predict the variance that certain factors account for, such as the age of the children in the sample, deprivation levels, program dosage, and experience the teacher had delivering DI programs. In line with preceding systematic reviews (e.g., Kinder, Kubina, & Marchand-Martella, 2005) and meta analyses (e.g., Borman, Hewes, Overman, & Brown, 2003) focusing on DI, Stockard et al's results indicated that DI programs elicit consistently positive results. Most effect sizes fell within the range considered medium ($d = 0.50$ to $d = 0.79$) to large ($d \geq 0.80$) in the context of the criteria used within general psychology literature (Cohen, 1988). However, in the context of educational interventions, most of the studies reported effects larger than $d = 0.25$, which reflects educational significance (Lipsey et al., 2012). These findings appeared to be robust with no impact of variables related to the nature of publication (year published and type of report/article), methodological approach, or sample. It is worth noting that these data reflect DI programs generally, but Stockard et al identified that studies evaluating mathematics programs, especially those aimed at older students (including the *Corrective Mathematics* program) were sparse.

Corrective Mathematics program

In Chapter 6, we used the *Corrective Mathematics* (CM) program to teach addition skills to children who attended a PRU. McGraw-Hill (2019) promote CM as a remedial DI program for children aged 8 to 18 years. Engelmann and Carnine (2005) organized the program into seven strategic modules, each progressing in difficulty: (1) addition, (2) subtraction, (3) multiplication, (4) division, (5) basic fractions, (6) fractions, decimals, and percentages, and (7) ratios and equations. Like other DI programs, CM comes with a placement test to help teachers assess which module and corresponding lesson each of their children should start on (Parsons, Marchand-Martella, Waldron-Soler, Martella, & Lignugaris-Kraft, 2004).

To the author's knowledge, only three published empirical studies have investigated the efficacy of the CM program. Glang, Singer, Cooley, and Tish (1991) reported a case study of an 8-year old boy, Thomas, who had sustained a traumatic brain injury after being hit by a vehicle. During the intervention, Thomas' teacher delivered lessons from both the CM program and the *Corrective Reading Comprehension* program (Engelman, Osborn, Haddox, & Hanner, 1978), to help develop his reasoning skills, ability to work out mathematics story problems, and develop his ability to recall arithmetic facts. The authors employed a multiple baseline design across content areas to evaluate the effectiveness of the program. As such, the teacher began instruction sequentially in each area after Thomas' baseline performance stabilized. Following instruction, Thomas' accuracy when answering story problems increased from 11.4% correct (baseline) to 91.25% correct. With regards to arithmetic facts, Thomas was able to answer an average of 6 facts per minute during the baseline phase and this increased to an average of 11.5 facts per minute during the instructional phase. However, the authors noted these data might be an underestimate of Thomas' performance due to his limited fine motor skills. If he had provided oral responses rather than writing, his performance would likely have increased further. These results reflect Thomas' progress over seven to eight DI sessions.

Sommers (1991) conducted an evaluation of DI programs in a school that had a program for at-risk children. The children in their sample were aged between 10 to 14 years and each scored below the 50th percentile on standardized mathematics and/or reading tests. None of the children qualified to access the special education resource room in their school but all were struggling to complete the content in their mainstream lessons. Following initial assessment, their teacher placed the children into groups of 5-10 to work on an appropriate DI program. They offered programs to support reading, mathematics, spelling, and expressive writing. The teacher delivered CM lessons to the children needing additional mathematics support from the following modules: multiplication; division; basic fractions; fractions, decimals, and percentages and; ratios and equations. Collecting data between 1985 and 1991, Sommers was able to establish the average age equivalence gains on standardized measures for each of the children receiving DI support. On average the children completed 7.6 months of instruction each academic year and their grade-equivalent scores on the standardized test improved by a mean of 1.01 years. These results provide evidence to suggest that the CM program helps children to learn more content in a reduced amount of time to close the attainment gap. However, it is unclear from this study the extent to which the literacy-based interventions impacted the children's results (particularly in the higher order modules that involve children to apply comprehension skills to written mathematics problems).

Finally, Parsons et al. (2004) assessed the efficacy of a peer-delivered CM program in a mainstream high school. Their sample consisted of ten learners who had previously failed the lowest-level mathematics class available at their school (entitled: integrated algebra), as well as nine peer tutors who had all passed a higher-level mathematics class (entitled: algebra II). A DI consultant trained the tutors in the use and delivery of the CM program so that they could run one-on-one sessions with the learners. Across a 10-week period, the learners and tutors completed five instructional sessions per week. On average, the learners made a standard score gain of 11.60 on the Woodcock-Johnson Psychoeducational Battery (WJ-R *ACH*) calculation subtest (i.e., they improved their score at pre-test by 2 standard deviations; *SD*) and a standard score gain of 5.80 on the applied problems subtest (i.e., they improved their score at pre-test by 0.89 *SD*). Interestingly, the peer tutors also made significant gains on the applied problems subtest across their engagement in the program; improving their standard score by 13.00 (1.30 *SD*). Collectively these studies suggest that the CM program can help children improve their mathematics skills, but further research is necessary to build a robust evidence base across other samples at risk of poor academic outcomes.

An instructional fluency approach

While DI helps children to acquire skills and knowledge, PT offers the tools for criterion referenced assessment and data-driven decision making. The strengths of these instructional technologies complement each other well. As such, educators have been combining PT and DI to help the learners that they work with make unprecedented academic gains (Binder & Watkins, 1990). To provide one applied example, Morningside Academy, in Seattle, integrate both technologies within their classrooms. Johnson (1997) explained that teachers at Morningside Academy allocate children with poor performance skills into classes that use DI programs to improve their reading, writing, and mathematics skills. Immediately following the completion of a DI lesson, the children continue to practice their new skills using PT methods. During this aspect of the session, children work on fluency-building tasks either independently, with their teacher, or with a peer tutor. This aims to help the children retain the skills that they have learnt, endure the skills for extended periods, and apply them to novel situations (i.e., achieve RESA). Within Chapter 6, we adopted a similar model whereby children completed timed practice sheets after each CM lesson to help assess and consolidate their skills. We subsequently used the SCC to help guide decisions about future practice.

The existing published literature exploring the instructional fluency approach (combining DI and PT) focuses largely on literacy skills. Case studies suggest that this approach can help children with autism learn to read in Icelandic (Adda Ragnarsdóttir, 2007), help adults to improve their reading of passages and nonsense words (Hulson-Jones, Hughes, Hastings, & Beverley, 2013) and help children holding a label of a learning disability to read sight words more fluently (Morrell, Morrell, & Kubina, 1995). When the sessions are delivered to children in small groups, the evidence suggests that children with special educational needs can improve their sentence construction (Datchuk, 2017). To the author's knowledge, little to no published research reports the combination of these techniques in relation to mathematics programs nor their application within a PRU.

Chapter 2 : **Methods and aims**

This chapter aims to provide some context and justification for the research methods and analysis techniques we adopted throughout Chapters 3, 4, 5, and 6. Collectively, the empirical chapters within this thesis help to address some wider research questions with regards to using PT and DI approaches within everyday classroom settings.

Generating an evidence-base

Thornicroft, Lempp, and Tansella (2011) conceptualized the journey towards establishing an evidence-base for medical and psychosocial interventions. Their model explains the key types of experimental research and some of the challenges that prevent the effective translation of evidence from early research to later implementation (see Figure 2.1). Within *phase 0*, researchers build theories, generate hypotheses, and conduct some small-scale laboratory studies to identify initial confounding variables. In some cases, researchers conduct these studies with animal populations which can affect the transfer of new knowledge to humans (t*ranslational block 1*). Within *phase 1,* researchers conduct studies that identify the key components of an intervention and begin to manualize it by testing its effects on human participants. Once an intervention has been clearly defined, we can begin some larger-scale exploratory studies (*phase* 2). Researchers use *efficacy* designs to evaluate the outcomes of an intervention when used with the target population. Typically, researchers use randomized controlled trials (RCTs) to evaluate the intervention under high levels of experimental control in ecological valid environments. These studies enable us to establish the outcomes of an intervention under optimal circumstances (i.e., when the environment is conducive to success) and also enable us to compare the intervention with an appropriate alternative (e.g., a competing intervention or treatment as usual group). *Translational block 2* indicates an interface between the outcomes of research from efficacy designs and those conducted under routine real-world conditions. *Effectiveness* studies enable researchers to establish whether an intervention can still yield positive results within the target population when the experimental control is reduced (*phase 3*). Sometimes the intended stakeholders do not adopt or maintain the intervention after the termination of a research study, despite the evidence-base for that intervention (*translational block 3*). The final phase, *phase 4*, encapsulates implementation research. This research identifies factors that interrupt the uptake of evidence-based interventions under everyday conditions in the real-world. This includes factors associated with the intention to implement the intervention, early implementation, and the persistence of implementation.

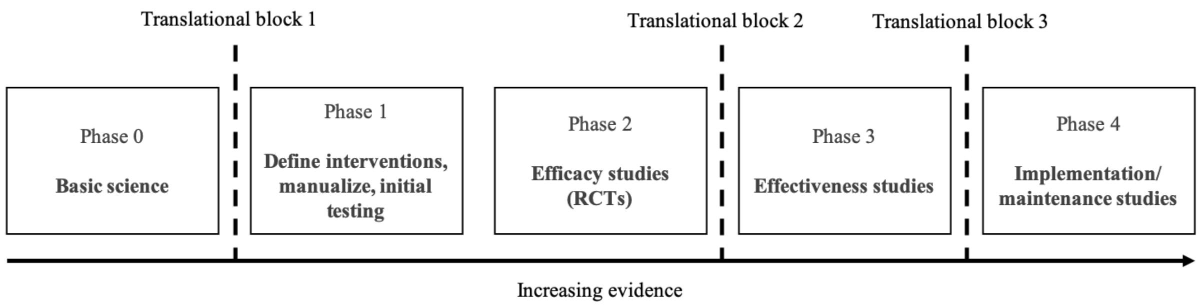

Figure 2.1. An adaptation of Thornicroft et al's (2011) model depicting the generation of an evidence-base for an intervention

In recent years, there has been a shift within education to ensure that evidence, rather than political will, informs policy (Morrison, 2001; Head, 2015). Hammersley (2005) explained that evidence-based practice within education is about integrating teachers' professional expertise with the best research evidence to improve the quality of educational practice. Professional development guidance often suggests that teachers should engage with the process and outputs of research in order to achieve this goal (Centre for Use of Research Evidence in Education; CUREE, 2011; Gorard, Huat See, & Siddiqui, 2020; OECD, 2016). When considering what evidence is necessary to inform practice, Nelson and Campbell (2017) argued that researchers and teachers should consider two questions: "what works?" and "what matters?". In order to answer these questions, we need to adopt different research methods depending on the phase of research and the context in which we are using the intervention.

Cluster-randomized controlled trial

Within Chapter 3, we employed a cluster-randomized controlled trial (c-RCT) to evaluate the benefits of teacher coaching (i.e., a researcher providing teachers with implementation support) on children's fluency of arithmetic facts during a SAFMEDS program. Assessing the impact of coaching was a timely research question in the context of the growing uptake of the SAFMEDS strategy in schools across North Wales (Tyler et al., 2019). As discussed in Chapter 1, the vast majority of existing SAFMEDS literature focuses on individual or small *N* delivery of a program by someone with prior expert knowledge of the strategy. Little research has explored how the strategy functions under the everyday conditions of a classroom, whereby a teacher or teaching assistant (TA) introduces and manages the program with the children in their class. At the beginning of our study, all members of teaching staff received the same SAFMEDS training before being randomly allocated to one of two research conditions (trial arms). Importantly, this study aimed to evaluate whether support after training improves fluency outcomes from a SAFMEDS program compared to receiving no support. The support model that we employed provided an option for delivering professional development and support to teachers at scale. Moreover, if effective, this coaching would provide teaching staff with the skills that they need to sustain a SAFMEDS program beyond the termination of the research study. Figure 2.2 is the graphic representation of the logic model underlying this study. This model outlines how our activities intended to improve the academic outcomes of the children within our sample.

Connolly, Biggart, Miller, O'Hare, and Thurston (2017) outlined the key characteristics of RCTs. In education, these trials assess whether a particular program/intervention is effective by comparing children's progress against a comparison (i.e., a group of children using a competing intervention) or a control group (i.e., a group of children who receive no additional intervention support and continue with classroom instruction as usual). During RCTs, a researcher randomly allocates children to a trial arm. As long as a study has enough participants, individual randomization aims to ensure that the arms of the trial have a similar profile of measured and potentially confounding baseline outcomes (Reich & Milstone, 2014). In theory, we can then attribute any differences between baseline and follow-up outcomes to the intervention itself; assuming that there are no threats to the validity of the randomization process or any substantial differentiated attrition of the sample throughout the study (Kraft, 2020). Subsequently, RCTs allow researchers to assess causality rather than correlation and many consider them to be the 'golden standard' research design within education (Connolly et al., 2017).

The distinguishing feature of a c-RCTs is that a researcher randomly allocates pre-existing groups, also known as clusters (e.g., schools), to each trial arm rather than individual participants (Piaggio et al., 2001). There are several reasons why a researcher may opt to use cluster randomization. For example, if the intervention occurs on a class-wide or school-wide level it may not be possible to randomly allocate individual children or teachers to a trial arm (Connolly et al., 2017). Clustering also helps to prevent contamination across trial arms. Borm, Melis, Teerenstra, and Peer (2005) explained that contamination can occur when individuals in different trial arms come together and influence each other. In the context of our study (Chapter 3), if schools shared a headteacher we treated them as one unit and allocated them to the same trial arm. In doing this, we hoped to prevent teachers receiving coaching from discussing their feedback and guidance with those allocated to receive no implementation support after training. During the allocation we also stratified schools based on their local authority area (county) and predominant teaching language (English or Welsh) to further anticipate some of the confounding variables that might have affected children's outcomes during the trial.

To assess the outcomes from c-RCTs, multilevel modelling enables us to analyze data that are hierarchically organized, such as children being enrolled in their corresponding schools. Connolly et al. (2017) explained that due to this hierarchy we can no longer treat each child as being independent from one another. Rather, statisticians assume that there is likely to be a correlation between characteristics of children that attend the same school. For example, schools tend to enroll children from the same neighborhoods (catchment area) and the children share the same classroom environment and instruction. Multilevel modelling is an increasingly popular analysis approach, which outperforms classical regression in terms of predictive accuracy. One feature of multilevel modelling that enhances its sophistication over classical regression is its ability to separate estimates for the predictive effects of covariates to produce group-level marginal means (Gelman, 2006). Multilevel models also produce an intraclass correlation coefficient (ICC), which is the proportion of variance in the outcome variable that can be explained by the levels (groupings) within the hierarchical model (Murray & Blitstein, 2003). It reflects the ratio of the between-cluster variance at that level of the hierarchy comparative to the total variance in the outcome variable.

A c-RCT was an appropriate research method for us to use to assess the need and impact associated with providing teachers with implementation support following SAFMEDS training. Cluster randomization enabled us to directly compare children's arithmetic fluency outcomes (using the Mathematical Fluency and Calculation Tests) and account for some of the naturally occurring confounding variables that may have otherwise affected the children's baseline outcomes or the success of the intervention. The associated analysis technique (multilevel modelling) provided some additional sophistication to the interpretation and accuracy of our results.

Effect sizes

In recent years, effect size has become a contemporary topic within education (Leonard & Woolcott, 2017). Effect size calculations allow us to quantify the magnitude of a phenomenon, by taking into account the average outcomes for different groups within a trial and the variance within the sample (e.g., *SD;* Kelley & Preacher, 2012). In education, researchers analyze outcome data that use arbitrary scales, such as academic achievement. Calculating an effect size from baseline and follow-up outcomes allows us to convert data onto a common scale that allows more meaningful comparisons between studies that used different measures. In turn, this enables researchers to compare the relative magnitude of different educational interventions (Kraft, 2020).

With leading educational organizations, such as the EEF, using effect size as a means of comparing interventions are becoming increasingly more cited within educational policy and decision making (Department for Education, 2016b; Higgins & Katsipataki, 2016). By default, in social sciences, researchers tend to evaluate the magnitude of their effect sizes using the thresholds proposed by Cohen (1988), whereby ≥ 0.20 *SD* is a small effect, ≥ 0.50 *SD* is a medium effect, and ≥ 0.80 *SD* is a large effect. However, Cohen himself recommended that researchers should only use these benchmarks when no better index is available (Kraft, 2020). Previous meta-analyses of field experiments within education have since demonstrated that effect sizes reported within intervention literature often reflect effects that are considered small or negatable by Cohen's original benchmarks (e.g., Cheung & Slavin, 2016; Lortie-Forgues, & Inglis, 2019).

We acknowledge that researchers, educators, and policy makers should be cautious when reporting and interpreting effect sizes in education. Simpson (2019) noted that effect size does not directly inform us about how educationally influential an intervention is. The calculation evaluates the trial as a whole rather than the intervention alone. Any changes to the fundamental elements of the trial (e.g., the population in which we recruit a sample, the chosen assessments) can influence the magnitude of the resulting effect size (Simpson, 2017). Kraft (2020) provided a new schema to outline several key trial elements that may significantly affect the magnitude of effect sizes (i.e., Cohen's *d*) within education studies. These include whether the effect sizes arise from correlational or causational studies; what, when, and how researchers assess outcomes, and; how researchers recruit their sample and the subsequent effects of selection bias. Moreover, the cost and scalability of an intervention may influence key stakeholder's views with regards to adoption of the intervention over alternative program with similar effects. That is, they may view the effect sizes from less costly interventions more favorably than similar effects from a more expensive alternative. Stakeholders may also favor the outcomes of research conducted with a larger-scale sample as it more accurately reflects how the intervention will operate at scale under real-world conditions.

In light of this information, Kraft (2020) proposed new empirical benchmarks and guidelines to reflect the effect sizes associated with student achievement. Researchers and educators should consider the effects of studies of preschool to Key Stage 5 educational interventions (3 to 18 years old) as small if the effect size is <0.05 *SD*, medium if the effect lies within the range 0.05 to ≤ 0.20 *SD*, and large if the effect is ≥ 0.20 *SD*. These benchmarks are contextually important when considering the academic outcomes in Chapter 3, whereby the comparisons between Cohen's (1988) benchmark and the effect sizes reported in the wider teacher coaching literature vary. We were able to report effect sizes for the analysis we presented in Chapter 3 due the large sample size of the study and the inclusion of a treatment as usual condition.

Qualitative research

Chapter 4 details two qualitative studies focusing on the social validity of the SAFMEDS strategy from the perspective of teachers and children who have used it in schools across North Wales. Specifically, the questions that we asked aimed to gain insight into teacher's experiences of implementation, perceived usefulness of the strategy, and factors that may affect children's engagement. By gaining insight into how the strategy works under everyday classroom conditions, we can better understand how to support schools with any unanticipated challenges that they may face (and pre-emptively troubleshoot, if appropriate). In turn this may contribute towards supporting teachers to persist with the implementation of the strategy and maintain/improve children's engagement. Furthermore, we can reflect upon the positive elements and consequences of the strategy to ensure that we continue to promote these factors in practice. To date no known published research has robustly examined the social validity that these key stakeholders associate with the SAFMEDS strategy. Figure 2.3 provides a graphical representation of the logic model underpinning the qualitative research we conducted and presented within Chapter 4.

Social validity refers to the perceived importance, acceptability, and viability of a program (Foster & Mash, 1999). Wolf (1978; as cited in Schwartz & Baer, 1991) formally introduced the term into the field of applied behavior analysis (ABA) to make researchers aware that non-acceptance of interventions by key stakeholders could lead them to reject the program. Surveys and interviews can be useful research tools to help us gather information about stakeholder's thoughts, experiences, and feelings. These methods enable us to access and understand behavior that is not always directly observable and quantifiable (Braun & Clarke, 2013).

To understand teacher's experiences of using the SAFMEDS strategy in schools, we disseminated an online survey (Chapter 4). Qualitative surveys provide a research method for gathering information about the diversity and experiences within a target population (Debski & Gruba, 1999). Van Selm and Jankowski (2006) highlighted some of the advantages of collecting research data using online surveys. For example, it is cheaper to deploy an online survey relative to the pen-and-paper alterative, it is more time efficient to disseminate an online survey, respondents can complete them at their own convenience, and online surveys have the potential to reach a larger number of respondents. Due to their in-built ability to preserve anonymity, online surveys may allow respondents, who may otherwise be difficult to reach, the space and confidence to share their experiences and opinions.

We devised the questions for our online survey to broadly capture teacher's/TA's experiences of using the SAFMEDS strategy within their school. These questions were direct in the sense that we asked teaching staff to identify what they perceived the benefits and challenges of the SAFMEDS strategy to be, but we tried to avoid priming any specific responses (for example, we did not ask "did any of the children in your group engage in cheating?", instead we asked "is there anything that the children do during a SAFMEDS session that you think hinders their performance?"). We believe that this helped us to collect a more holistic view of how the strategy functions under the everyday conditions of a classroom. If the teachers identified any unexpected consequences of using the strategy or factors that affected implementation, then this may provide the basis for more direct questioning in future qualitative studies.

Patton (2002) defined an interview as a series of open-ended questions that gain insight into people's experiences, perceptions, opinions, feelings, and/or knowledge. Within Chapter 4, the researcher conducted one-on-one semi-structured interviews with children who had used the SAFMEDS strategy with a teacher/TA in school. By nature, semi-structured interviews begin with pre-defined questions, but the interviewer can ask additional un-scripted follow-up questions (Patton, 2002). Prompting for elaboration during a semi structured interviews can be an invaluable tool for ensuring the reliability of data. It allows the researcher to ask for clarification of any interesting and relevant issues raised by the interviewee (Hutchinson & Wilson, 1992). It also allows the researcher to clarify any inconsistencies within an interviewee's responses (Barriball & While, 1994).

We opted to use the semi-structured interview approach with the children in our sample for two main reasons. Firstly, we wanted to observe them engaging in the SAFMEDS strategy to assess for fidelity. This may have led to further questions if they had adapted the procedure in any way (e.g., if they asked the researcher to hold and direct the cards for them). The researcher may also have decided that the child did not follow the strategy closely enough (compared to the manualized program) to continue to interview. Secondly, the children included in the sample within Chapter 4 were aged between 6 and 10 years old. We believed that using semi-structured interviews would lead to the richest and most detailed responses for us to be able to analyze. Semi-structured interviews are somewhat reactive, so this gave the researcher the opportunity to rephrase the question if a child did not demonstrate understanding, ask additional questions for clarity or expansion, and explore additional relevant information as the interview unfolded. Kortesluoma, Hentinen, and Nikkonen (2003) explained that using interviews with children can be a favorable method of collecting qualitative data over surveys because it allows them to have space and time to discuss their views and experiences, without limiting their responses to a narrow range of categories or Likert scales. Their responses are also not bound by their written comprehension abilities.

Researchers can interpret meaning in qualitative data by identifying themes across participant's responses (Tolich & Davidson, 2003). Ritchie and Lewis (2003) explained that there are no clearly agreed rules for analyzing qualitative data. Rather, analysis approaches vary based on their epistemological assumptions, the nature of enquiry, and the main focus of the analytical process. Thematic analysis is a method that researchers can use to systematically identify and organize qualitative data into themes (patterns of meaning) across the dataset. Crucially, thematic analysis allows researchers to make sense of a targeted *group* of people's experiences by identifying commonalities as opposed to analyzing data on a person-by-person basis (Braun & Clarke, 2012). The form of thematic analysis that we used within Chapter 4 focused on using an inductive (bottom-up) approach to data coding. This means that the codes and themes emerge from the content of the transcripts themselves. This is opposed to a deductive (top-down) approach, whereby a researcher would align the data extracts with pre-conceived concepts, ideas, or theories (Braun & Clarke, 2006). By following an inducive approach we hoped to report a more holistic overview of the strengths and challenges of using the SAFMEDS strategy as a learning tool, as this is something lacking in the current literature.

Braun and Clarke (2006) provided clear and comprehensive guidance to support researchers through the phases of thematic analysis. First, a researcher should familiarize themselves with the participant's responses. They should transcribe the data (if necessary) and then read through the completed transcripts several times before beginning analysis. This is to ensure that they become immersed in the data and become familiar with the depth and breadth of the content. Through repeated iterations of reading the transcripts the researcher should become increasingly more active in the way they respond to it. That is, they should start searching for meaning and patterns within the dataset. As the researcher progresses through this immersion phase, they should start to take notes about the data that they can go back to during later phases of analysis (i.e., note any interesting and salient ideas that they notice).

The second phase of thematic analysis involves generating some initial codes from the data. Codes reflect the most basic element of raw data that the researcher can assess in a meaningful way in relation to the phenomena of interest (Boyatzis, 1998). Braun and Clarke

(2006) advised that researchers code for as many potential themes as possible, as they will not know what they might consider as interesting and relevant until later on in the process. They also advised that researchers should keep some surrounding text in their extracts to ensure that context is not lost in later phases.

During the subsequent phase, the researcher should focus their analysis to look at broader level themes, rather than codes. Braun and Clarke (2006) recommended using visual aids to help organize these ideas, such as writing each code on a piece of paper and rearranging them into theme-piles. This process should lead the researcher to start drawing a thematic map that depicts the relationship between codes, between themes, and between sub-themes. Researchers should end this phase with candidate themes and the corresponding extracts of data that they have coded in relation to them. Whilst the researcher will have a sense of the significance of each theme, Braun and Clarke warned against excluding anything at this phase (even if there are miscellaneous codes) as they will not know whether the themes are robust until they evaluate all of the extracts in detail.

When reviewing themes, it will become clear to the researcher that some of the candidate themes do not hold (e.g., there is not enough data to support them or they are too diverse), whilst other candidate themes may need to merge because they reflect the same thing. Braun and Clarke (2006) explained that data within themes should come together in a meaningful way and there should be clear distinctions between separate themes. There are two levels to the refining and reviewing process. Within level one, the researcher should review the themes at the level of the data extracts. The theme should capture the contours of the coded data. Within level two, the researcher should review the themes at the level of the entire dataset. They should consider the validity of individual themes and whether the thematic map accurately captures the data as a whole, if not they can redefine and recode the data until they achieve this. By the end of this phase, the researcher should have a good idea of what the themes are, how they fit together, and the story that the data tells.

Once the researcher has a completed thematic map they can then name and define each theme (Braun & Clarke, 2006). The researcher should identify the essence of what each theme is about without each theme being too diverse and complex. To do this, they should go back to the data extracts for each theme and organize them into a coherent account that follows a narrative. During this process, the researcher should not simply paraphrase the data extracts, rather identify what is interesting about them and why. The name of each theme should be concise and immediately inform the reader what the theme is about.

The final phase of thematic analysis consists of producing the report/publication. Braun and Clarke (2006) outlined that researchers should aim to tell the story of their data in a way that convinces the reader of the merit and validity of their analysis. During the write-up, researchers must provide sufficient evidence to support each theme in the form of the original data extracts. When choosing extracts to present, Braun and Clarke advised that researchers select quotes of vivid and simplistic examples that reflect the essence of each theme. These quotes should support the narrative to help the researcher to make an argument in relation to their original research question(s), rather than being purely descriptive.

The two studies that we present in Chapter 4 aimed to help us gain insight into teacher's experiences of implementing the SAFMEDS strategy under the day-to-day conditions of the classroom, perceived usefulness of the strategy, and factors that may affect children's engagement with the strategy. By using open ended questions and an inductive analysis approach, we hoped to explore the social validity that these key stakeholders associate with the strategy and identify factors that may impact implementation of the strategy over time.

We decided to present and analyze the data within Chapter 4 as two separate studies for a few core reasons. Firstly, we asked the children and teachers different questions; whilst some of the emerging codes might have overlapped, there was no guarantee upon data collection and analysis that this would be the case. Secondly, we used two different research methods to generate the data across the two studies. We expected teachers' responses to the survey to be more in depth and unprompted, whilst the semi-structured interviews required more prompting and exploration during the data collection process. Finally, by combining the datasets we were worried that some key ideas would lose their saliency and therefore not accurately reflect the experiences of the two discrete samples (teachers and children).

Impact reports

XLP are a youth-charity based in London, United Kingdom. As well as supporting social issues such as unemployment and preventing gang involvement, XLP provide resources to try and remediate and prevent educational failure (XLP, n.d.a). The charity contacted the research team at Bangor University after hearing about how the SAFMEDS strategy can help support children's fluency of basic academic skills. After an initial consultation and some training, we worked collaboratively over two years on the XLP literacy and numeracy project (XL-LAN). Chapter 5 outlines the process underlying this collaboration—from sharing and designing SAFMEDS packs, to the analysis of project data, and generation of the impact report. This report aimed to provide some exploratory descriptive statistics to better understand how this demographic of children *typically* progress on a session-by-session basis when using SAFMEDS to practice and assess their skills.

Close-to-practice research helps to generate knowledge that is relevant to key stakeholders and addresses practice concerns. This type of research helps to develop a research culture within organizations by supporting individuals and/or teams to think critically about gaps in their knowledge and services, use appropriate methods to elicit change, and evidence/evaluate outcomes (Cooke, 2005). The British Educational Research Association (BERA) has identified close-to-practice research as an area of interest, particularly in the context of teacher education and professional development. The association have also identified that such educational research should focus on addressing problems in practice and encouraging researchers to develop working partnerships with key stakeholders within education (Wyse, Brown, Oliver, & Poblete, 2018).

Chapter 6 illustrates an example of how schools, charities, and researchers can work collaboratively to foster a closer-to-practice project; specifically focusing on raising children's fluency of basic literacy and mathematics/numeracy skills. This project did not follow a typical pre-and-post-test design, rather the quantitative results reflected the data generated through the SAFMEDS timings themselves. We used these data to establish the average progress that children can make using the strategy and how this may differ depending on whether they are focusing on literacy or mathematics skills.

Moreover, we captured some qualitative data from the children who participated in the project to gain insight into their experiences and opinions of using the strategy to support their learning. The age of children involved in the XL-LAN project ranged from 7 to16 years old. In order to promote consistency and inclusivity in the research methods that we used, we opted to conduct semi-structured interviews (as outlined and justified earlier within this chapter). Whilst we appreciated that conducting a fidelity check before the interviews would have been beneficial to enhance validity, there were some additional timing and practical constraints attached to this project (e.g., the time XLP staff had allocated to be in schools, the time they could spend with the children before they needed to go to their next lesson, having to conduct some of the interviews in small groups rather than one-on-one). However, the project coordinators at XLP had worked closely with the schools involved in the XL-LAN project to provide support with implementation. As such, if there were any major concerns with how the children engaged with the strategy the staff would have addressed these earlier in the school year.

Together, the data we presented within the impact report served to support two outcomes. First, to provide XLP and the schools with an evaluation of the XL-LAN project and some recommendations for future practice. XLP needed this information to provide to the charities and external bodies that had provided funding for the project. The data and recommendations aimed to help inform future proposals for extensions to the service that XLP can offer to schools and the communities that they work with. Second, the quantitative data provided us, as researchers, with some additional insight into how groups of children may progress using the SAFMEDS strategy to support their learning, whilst the qualitative data provided an extension to what we had discovered within Chapter 4. The qualitative data also helped develop our understanding of how children attending secondary schools (ages 11-16 years) perceived the strategy.

Small *N* design

In the final empirical chapter, Chapter 6, we present a small *N* pilot study focusing on using an instructional fluency approach to teach addition skills to children attending a PRU. There are no known published studies that have reported the use of DI and/or PT to support children who attend these provisions. As such, we piloted the use of the *Corrective Mathematics* (CM) program to support a small sample ($N = 5$) of children to develop their addition skills and complemented the curriculum with paper-based fluency timings (adhering to PT methods). We hoped that the outcomes of this efficacy study would form a foundation for future research investigating the use of these methods to raise attainment within PRUs. Refer to Figure 2.4 to see the logic model underpinning this study.

Graham, Karmarkar, and Ottenbacher (2012) explained that small *N* designs require several observations of participants before, during, and after an intervention period.

Conventionally, researchers visually inspect the graphed and tabulated data within and across phases to draw conclusions about the results. Small *N* studies rely on replication rather than large sample statistics to establish reliability and enhance generalizability to other people and settings. The Small *N* approach is well established within ABA literature; including assessing the effects of PT (for examples, see Casey et al., 2003; Chiesa & Robertson, 2000; McDowell & Keenan, 2001) and the instructional fluency approach (for examples, see Blackwell, Stookey, & McLaughlin, 1996; Gregory, McLaughlin, Weber, & Stookey, 2005) to teach academic skills. Small *N* research can be particularly useful when exploring individualized interventions and working with children with additional learning needs. They are also a favorable research method if researchers lack funding and resources to recruit larger samples or when the population comprises of a small sample size, so it is not viable to recruit a large sample (Barnett et al., 2012).

PT methods lend themselves well to Small *N* designs due to the use of the SCC to monitor progress on a child-by-child basis. As discussed within Chapter 1, the SCC provides a way of visualizing fluency data. Practitioners can assess celeration trends (i.e., the trajectory of the data at a certain portion of the SCC) and the variability of the data (also referred to as bounce) via visual analysis. Moreover, the addition of phase change lines on the chart allow a practitioner to see whether the change they have implemented had the desirable effect on the dependent variable or not (White & Neely, 2012). An additional advantage of continuous analysis, such as this, is that practitioners are able to identify and control confounding variables and thus stabilize the dependent variable. For example, the practitioner could manipulate something in the environment instantaneously to determine whether it, or another confounding variable, is responsible for an observed fluctuation in the dependent variable (Saville & Buskist, 2003).

We opted to use a Small *N* design for this study primarily due to the diversity of the sample. As discussed in Chapter 1, PRUs accommodate children on either a singular or dual-registry basis, with a large proportion of children also holding a statement for at least one additional learning need (Department for Children, Schools and Families, 2008; Parliament UK, 2018). This raised challenges in terms of creating a control group based on a matched pairs design, as children attended different mainstream schools for varying hours per week, they were different ages (so have been exposed to different quantities of the curriculum), and due to their additional learning need(s) may have faced different challenges accessing the material. It is also worth noting that due to the timescale available to conduct the study, the researcher was also unable explore the possibility of using the children as their own controls (e.g., employing an AB design).

At the beginning of our study (Chapter 6), we assessed the children's baseline mathematics ability. By using published standardized academic tests (the Test of Early Mathematics Ability and the Wide Range Achievement Test) we hoped to gain a broad overview of the children's skill set. In doing so, we were able to better understand where they had deficits in their knowledge and, if appropriate, use this information to tailor subsequent instruction during the program. By re-administering these tests after the intervention period, we were also able to provide some additional data regarding each child's progress across the intervention.

Context

- Wales placed significantly lower than other developed countries in the 2009 and 2015 PISA assessments for mathematics (OECD; 2010, 2016).
- This prompted Welsh Government to initiate an education reform to help improve student outcomes; including the adoption of more evidence-based strategies in schools (Furlong, 2015).
- Children who are eFSM are proportionally more likely to underperform across the curriculum.
- The EEF (2019) suggest that schools invest in follow-up support after training for educational interventions to help yield the best results.
- The existing published literature on the SAFMEDS strategy suggest that both teachers and researchers can support learners to develop fluency. However, it is still unclear what role a researcher can contribute to learner outcomes in teacher led contexts.

Aims

- Assess the impact a researcher can offer to a teacher led mathematics program, through a coaching model (three 1-hour visits and email contact).
- Improve children's ability to perform arithmetic fluently.
- Evaluate a coaching model that would offer a more cost-effective and feasible alternative to embedding a researcher in schools to run and maintain a SAFMEDS intervention.

Rationale/theory of action

- Initial training to implement a SAFMEDS program would improve teacher's pedagogical knowledge of the strategy and the associated data-driven practice.
- The in-house coaching would offer individualized feedback and support to improve teacher's fidelity of implementation.

Assumptions

- SAFMEDS is a practice and assessment strategy. Children can use the strategy to practice skills to fluency and identify any skill deficits.
- Visualizing data on an SCC makes it possible to monitor children's progress on each SAFMEDS skill. If they are not making desired progress towards mastery, additional instruction, change in environment, or change in materials might be necessary.
- The in-house coaching would provide feedback with teachers to support and/or improve their fidelity of implementation

Inputs

- ESRC DTP grant (academic input).
- Supervisory team
- SAFMEDS expert to coach
- Statistics and randomization support.
- Teacher/TA time to run and sustain intervention.
- Child engagement.

Activities

- Gain ethical approval.
- Distribution of opt-out consent forms.
- Teacher training session.
- Training blind assessors to administer the assessments.
- Teachers run 3 SAFMEDS sessions (approx. 20-minutes each) per week in their school.
- Teachers to plot and review children's data regularly.
- Three 1-hour support visits and email support for each school in the support trial arm.
- Troubleshooting of technical issues for schools in both trial arms.
- Administration and analysis of Baseline and follow-up assessments.
- Write an article for peer review.

Outputs

- Teachers and TAs trained to use the SAFMEDS strategy in their schools.
- Analysis of MFaCTs data assessing fluency progress.
- 1 published article in a peer-reviewed journal.
- Dissemination of findings at conferences.
- An impact report provided to schools outlining the project and findings.

Participants

- Teachers and TAs who deliver the SAFMEDS sessions.
- Children enrolled on the SAFMEDS program.
- Research team – PhD student, supervisory team, analysis support.
- Independent statistician to conduct the randomization.
- Assessors – each blind to the trial arm each school had been allocated to.

Outcomes/impact

- **Short (2017-18):**
  - Increased knowledge and use of the SAFMEDS strategy in school across North Wales.
  - Children working towards developing arithmetic fluency.
  - Establish the impact of the current coaching model on student outcomes.
- **Medium (2018-19):**
  - Teachers able to continue using the SAFMEDS strategy after the termination of the research study to support fluency development.
  - Research team to offer follow-up or "refresher" training to schools across North Wales.
  - Initial dissemination of findings (e.g., conference presentations/posters).
- **Long (2019 +):**
  - Sustained use of the SAFMEDS strategy in schools, with high levels of implementation fidelity.
  - Children continue to master basic skills to support them across the curriculum.
  - Continual training opportunities.
  - Further dissemination of findings (e.g., peer reviewed publication).
  - Further exploration of factors associated with implementation (e,g, social validity, challenges, effect of coaching on implementation fidelity).

Figure 2.2. The logic model underpinning our c-RCT (Chapter 3)

Context

- The existing SAFMEDS literature suggests that the strategy can help learners to develop skills to fluency and offers a level of versatility (e.g., scalability, use with diverse learner populations, to practice skills across the curriculum).
- Most of the peer-reviewed research reports the quantitative fluency outcomes of SAFMEDS programs, with few reporting anecdotal findings relating to social validity (e.g., Hunter et al., 2016).

Aims

- To conduct a more robust evaluation of stakeholder's views on the SAFMEDS strategy.
- Gain insight into the advantages and challenges that educators (e.g., teachers, TAs) associate with using the strategy in their classrooms; with a specific focus on outcomes and implementation.
- Gain insight into children's experiences of using the SAFMEDS strategy; with a specific focus on factors that may affect engagement.

Assumptions

- Teachers, TAs, and children who participate in this research will have directly used the strategy within their school regularly, for a sustained period during at least one school term.
- We assume that the implementation fidelity check is an accurate representation of how children typically engage with the SAFMEDS strategy in their school.

Inputs

- ESRC DTP grant (academic input).
- Research student (PhD)
- Supervisory team
- Teacher and TA time to complete the online survey
- Children's time to participate in the interviews.

Activities

- Gain ethical approval.
- Create and deploy an online survey to collect teacher's views and experiences.
- Devise interview questions to capture children's views (these need to be accessible to children aged 6-years +).
- Distribution of opt-in consent forms to interview the children.
- Conduct interviews with children.
- Transcribe interview recordings.
- Analyze study data using thematic analysis.
- Write an article for peer review.

Outputs

- Analysis of online survey responses and interview transcripts, reflecting teacher are children's experiences of the SAFMEDS strategy.
- 1 published article in a peer-reviewed journal.
- Dissemination of findings at conferences.
- Representation of these themes in future training for the SAFMEDS strategy in North Wales.

Participants

- Teachers and TAs who deliver the SAFMEDS sessions.
- Children enrolled on the SAFMEDS program.
- Research team – PhD student and supervisory team.

Outcomes/impact

- **Short (2018-19):**
    - Insight into teacher's experiences of implementing the SAFMEDS strategy in their schools, as well as what they perceive the benefits of the strategy to be.
    - Insight into children's experiences of using the SAFMEDS strategy in school, with a specific focus on factors that might affect engagement.
- **Medium (2019-20):**
    - Inform teachers of the results of this research during initial training of the SAFMEDS strategy (to help highlight what works and ways to overcome challenges).
    - Initial dissemination of findings (e.g., conference presentations/posters).
- **Long (2020 +):**
    - Sustained use of the SAFMEDS strategy in schools, leading to mastery of basic skills.
    - Continual training opportunities.
    - Further dissemination of findings (e.g., peer reviewed publication).
    - Further exploration of factors associated with implementation (e.g., investigating how crucial the challenging elements are to the effects of the intervention in the classroom).

Figure 2.3. The logic model underpinning our qualitative research (Chapter 4)

Context

- Estyn (2017) revealed that four in ten pupil referral units (PRUs) did not support children to develop the numeracy skills they needed to support their learning across the curriculum.
- The incoming Curriculum for Wales places an emphasis on making mathematics and numeracy experiences engaging and accessible for all children (Donaldson, 2015).
- The Curriculum for Wales 2022 guidance prescribes that children need to be able to perform the four basic arithmetic operations (addition, subtraction, multiplication, and division) fluently and understand the relationship between them (Welsh Government, 2019).
- Direct Instruction (DI) programs comprise of scripted lessons that teach learners skills in a sequential, explicit, and scaffolded order (Kinder & Carnine, 1991).
- Precision Teaching (PT) is a method of assessment that Leads to data-driven decisions about subsequent instruction (Johnson & Street, 2013).
- Slocum (1993) argued that combining DI and PT creates a superior technology to help learners master key concepts.

Aims

- Assess the effects of using an instructional fluency approach (i.e., the combination of DI and PT) to teach and build fluency of addition skills in a PRU.

Rationale/theory of action

- The DI lessons will help learners to acquire new knowledge and skills.
- The PT timings will help learners to practice these newly acquired skills to fluency.
- By charting the learner's data on a Standard Celeration Chart (SCC), the researcher will be able to make subsequent data-informed decisions about instruction.

Assumptions

- The *Corrective Mathematics* (CM) placement test identifies an appropriate skill level for each child (e.g., they have the appropriate prerequisite skills to be able to follow Lesson 1).
- Children's performance on the baseline and follow-up assessments are an accurate reflection of their test taking ability and skillset.

Inputs

- ESRC DTP grant (academic input).
- Supervisory team
- PhD student
- Child engagement.
- Blind assessor to administer baseline and follow-up assessments.

Activities

- Apply for and gain ethical approval for the research study.
- Distribute opt-in consent forms.
- Recruit a PRU to host the research project through head teacher meeting and email correspondence.
- Administration and analysis of baseline and follow up assessments (WRAT-4, TEMA-3, and CM placement test).
- Training blind assessor to administer the assessments.
- Researcher to facilitate the 1:1 instructional fluency sessions 3 times per week.

Outputs

- Pre-post analysis of WRAT-4. TEMA-3, and CM placement test data.
- SCCs for the children who completed the intervention, making it possible to visualize their fluency and skill progress across the intervention period.
- 1 peer reviewed article in a peer-reviewed journal.
- Dissemination of findings at conferences.
- An impact report provided to the headteacher of the PRU outlining project and findings.

Participants

- Input of the head teacher to help arrange logistics and support the running of the project.
- Researcher (PhD student) to deliver the instructional fluency sessions.
- Children ($N \approx 5$) attending the PRU to participate in the sessions.
- Supervisory team.
- Assessors – each blind to the aims of the study.

Outcomes/impact

- **Short (2019-20):**
  - Increased knowledge of how the instructional fluency approach may function in a PRU.
  - Supporting the children who participate to improve their academic skills (with specific focus on addition).
  - Insight into the social validity children associate with using this method to support their skill development.
  - Increased understanding of the feasibility and acceptance of the intervention in a PRU-context.
- **Medium (2020-21):**
  - Generation of discussion with key stakeholders regarding the adoption of an instructional fluency intervention in PRUs.
  - This may lead to further investigation of the intervention and the generation of more research questions (e.g., can we scale-up the program for group delivery in PRUs? Can TAs deliver sessions to high levels of fidelity following basic training?).
  - Dissemination of findings (conference representation and a peer-reviewed paper).
- **Long (2021 +):**
  - Increased uptake of the instructional fluency approach to teach mathematics skills to children who attend PRUs.
  - Professional development opportunities for teachers and TAs to support implementation of the instructional fluency approach.
  - Children continue to master key numeracy skills to support their learning across the curriculum.

Figure 2.4. The logic model underpinning our small N pilot research (Chapter 6)

Chapter 3 : Implementation support improves the fluency outcomes of a fluency-based mathematics strategy: A cluster randomized controlled trial

Owen, K., Hunter, S. H., Watkins, R. C., Payne, J., Bailey, T., Gray, C., Hastings, R. P., & Hughes, J. C. (under review). *Implementation support improves outcomes of a fluency-based mathematics strategy: A cluster-randomized controlled trial.* Manuscript submitted for publication.

Preface

As outlined in Figure 2.2. there were several contextual factors that led us to investigate how researchers can support teachers to implement the SAFMEDS strategy in schools. With recent reform efforts in Wales to improve the use of evidence-base practice in schools (Furlong, 2015), it is important that we understand how these interventions work under the day-to-day conditions of the classroom. In order to promote professional development and high-quality implementation, leading educational organizations such as the EEF (2019) suggest that schools invest in follow-on support after attending training. Over recent years, a growing number of teachers in North Wales have received training to implement the SAFMEDS strategy in their schools to support the mathematics and numeracy development of the children that they work with (Tyler et al., 2019). Despite this, the vast majority of the published papers reporting the effects of the SAFMEDS strategy use efficacious designs and focus on researcher led implementation (e.g., Casey et al., 2003; Hunter et al., 2016). Our paper aims to provide insight into the follow-on support a researcher can offer after teacher initial training. This aimed to help us understand whether this support is necessary in the context of a teacher led SAFMEDS program.

This chapter contains the manuscript of the paper that we have submitted for publication. At the time of the submission of this thesis, this paper has undergone peer review and the journal have conditionally accepted it. Here we present the latest revision, which we have edited based on the feedback from the reviewers. We have also edited the formatting in line with APA guidelines (including placing figures and tables in the main body of the text). In Chapter 7 we discuss the implications, applications, and limitations of this research in more detail.

Abstract

The Say-All-Fast-Minute-Every-Day-Shuffled (SAFMEDS) strategy promotes fast and accurate recall. The existing literature suggests that the strategy can help learners improve academic outcomes. Through a cluster randomized controlled trial, we assessed the impact of implementation support on children's mathematics outcomes during a teacher led SAFMEDS intervention. Following training and prior to baseline assessments, we randomly allocated schools to receive either no ($n = 31$) or ongoing ($n = 33$) support from a researcher. Support consisted of three in-situ visits and email contact. Assessors remained blind to the condition of the schools throughout. We analyzed the outcomes of children ($n_{\text{Support}} = 294$, $n_{\text{NoSupport}} = 281$) using a multi-level mixed effects model; accounting for the children nested within schools. The results suggest that implementation support has a small effect on children's fluency of arithmetic facts (Mathematics Fluency and Calculation Tests (MFaCTs): Grades 1-2, $d = 0.23$, 95% CI: 0.06 to 0.40; MFaCTs: Grades 3-5, $d = 0.25$, 95% CI: 0.08 to 0.42). These results are larger than the average effect sizes reported within professional development literature that apply coaching elements to mathematics programs.

Evidence-based practice has the potential to generate effective educational programs and promote positive outcomes for students. Yet, Klingner, Boardman, and McMaster (2013) outlined that educators are less likely to adopt and sustain these practices at scale without strategic and systematic support from researchers. Implementation science suggests that researchers first conduct efficacy studies on a small scale to validate an intervention before implementing them at a larger scale under real-world (day-to-day) conditions. The quality of implementation during the latter phase may bound the benefits of evidence-based programs in school environments (Cook & Odom, 2013). Implementation fidelity refers to the extent to which someone implements a program according to the original and intended design (Lee, Penfield, & Maerten-Rivera, 2009). Durlak and DuPre (2008) found that educators who do not specialize in research (e.g., teachers) often do not implement an intervention to 100% fidelity under the real-world conditions of a classroom. They also found that low-quality implementation of evidence-based interventions results in smaller effect sizes on outcome variables; including those linked to student achievement. This highlights the importance of identifying effective implementation support models to ensure that teachers are able to elicit desired and intended outcomes from evidence-based educational interventions.

Training can be an effective way of helping teachers to develop conceptual understanding of interventions but alone may not yield sufficient changes in practice (Education Endowment Foundation, 2019). Coaching teachers offers a lever for improving the quality of implementation by supporting them to translate knowledge into classroom practice (Kraft, Blazar, & Hogan, 2018). Sailors and Shanklin (2010) used the term coaching to describe a process of sustained school-based support from a knowledgeable individual. Coaches model research-driven interventions and work with teachers to explore how they can use the strategies with their own students. Coaching programs can take a variety of forms, but generally consist of one-on-one interactions between a coach and a teacher. These interactions provide a platform for teachers to receive individualized feedback based on their professional development (PD) needs (Fletcher & Mullen, 2012).

Following a meta-analysis of 60 studies, Kraft et al. (2018) found that teachers often receive coaching in conjunction with additional treatment elements (i.e., in 90% of the reviewed studies teachers received coaching alongside group training, instructional content, and/or video resources). Their analysis revealed a pooled effect size of +0.18 standard deviations (*SD*) relating to the effect of these programs on student achievement and +0.49 *SD* relating to teachers' instructional practices. In their theory of action, Kraft et al. outlined that training sessions help improve teacher pedagogical and content knowledge. This knowledge, alongside coaching and the availability of relevant materials, positively influences teaching behavior. As a result, teachers implement higher-quality teaching practices and are better able to identify and use strategies that support student outcomes. However, it is worth noting that most of the interventions that met Kraft et al's inclusion criteria focused on applying these practices to literacy and content-based interventions; with only two studies reporting the outcomes of mathematics programs. Moreover, Kraft et al's. analysis revealed that the effects on student outcomes from larger-scale effectiveness trials were smaller (+0.10 *SD*) than those employing smaller-scale efficacy designs (+0.28 *SD*). Whilst coaching might be a valuable tool, research is still needed to disentangle the effects of coaching from additional treatment elements and to establish the effects of using coaching programs at scale.

In a complementary meta-analysis of 95 studies, Lynch, Hill, Gonzalez, and Pollard (2019) reviewed mathematics and science interventions supported by PD and/or curriculum materials. Lynch et al. defined PD as a set of experiences that intend to affect change in teacher- and classroom-level phenomena. They too highlighted that PD programs can be, and often are, multifaceted. As such, their inclusion criteria focused on the number of hours a teacher spent experiencing PD; the focus on improving knowledge of content, pedagogy, and/or use of curriculum materials; as well as the format of the program (e.g., one-on-one coaching, summer workshops, online learning). Curriculum materials are instructional practices, guided by activities and text within the program itself. Their review identified that 22% of studies focused on PD alone, whilst 75% used PD in combination with curriculum materials. Overall, they found PD programs to be effective. However, only 20% of the PD programs included a coaching element. There was no evidence that coaching elements added value in terms of outcomes but neither did they reduce intervention's effectiveness. The vast majority of included studies with a coaching element were multi-component programs. With few published studies reporting the outcomes of coaching as a standalone PD format to support mathematics interventions, further research is needed.

SAFMEDS overview and prior research

The current study is set within the context of North Wales, United Kingdom. Following disappointing results in the internationally comparative Programme of International Student Assessment tests in 2009 (OECD, 2010), the Welsh Government identified a need to raise educational standards in their schools. In recent years, education policymakers in Wales have focused on improving the use of evidence-based practice within education (Furlong, 2015; OECD, 2016). As a result, an increasing number of teachers in North Wales are using the Say-All-Fast-Minute-Every-Day-Shuffled (SAFMEDS) strategy in their classrooms to improve children's fluency of basic mathematics skills (Tyler et al., 2019).

Traditionally within educational practice, teachers deem children to have mastered skills if they are able to perform them to a level of 90-100% accuracy (Fuller & Fienup, 2018). Binder, Haughton, and Bateman (2002) argued that a percentage correct criterion is too simplistic—being accurate is necessary, but not sufficient to demonstrate mastery of content. If children practice skills beyond mastery, they will be able to develop fluency (the combination of accuracy *and* speed). Adding a dimension of time into assessment provides more detail about performance and can more accurately predict whether children will be able to retain, apply, and generalize learned skills (Binder, 1996; Johnson & Street, 2012). SAFMEDS is a practice and assessment strategy that applies the principles of precision teaching (PT) to help children develop their skills to fluency (Lindsley, 1995).

Kubina and Yurich (2012) described PT as a system for defining, measuring, recording, and analyzing teaching effectiveness on a child-by-child basis. To achieve this, teachers must reflect upon children's learning regularly, and use these data to make subsequent decisions about their teaching approach (Lindsley, 1995). A child engages with the SAFMEDS strategy using a deck of flashcards, with a question or statement on the front and the corresponding correct answer on the back. They read the front of the card silently before vocalizing the answer (Quigley, Peterson, Frieder, & Peck, 2018). During each 1-minute timing, they aim to get through as many cards as possible, whilst separating their correct responses from their "not yets" (Cihon, Strutz, & Eshleman, 2012). The child then plots their best score from the session on a Standard Celeration Chart (SCC), which develops a learning picture over time (for more details see Lindsley, 1995). Learning pictures enable teachers, children, and/or other practitioners to decide if additional support is necessary. For example, if a learning picture shows that the number of cards a child can answer correctly in 1-minute has plateaued over several consecutive sessions, an intuitive approach might be to assess if the child has mastered all of the necessary prerequisite skills, and if not, to ensure that they do so (Johnson & Street, 2012).

The SAFMEDS strategy has clear utility within schools, with a growing quantitative evidence-base suggesting some of the associated advantages. For example, practitioners can adapt the strategy to implement it on a one-on-one basis (e.g., Cunningham, McLaughlin, & Weber, 2012), with small groups of children (e.g., Beverley, Hughes, & Hastings, 2018), or class-wide (Hunter, Beverley, Parkinson, & Hughes, 2016). The strategy also has evidence to support its effectiveness amongst different populations including learners attending mainstream classes and children with additional learning needs (e.g., Casey, McLaughlin, Weber, & Everson, 2003; Greene, Mc Tiernan, & Holloway, 2018; Kubina, Ward, & Mozzoni, 2000).

Much of the available literature on the SAFMEDS strategy documents small *N* and case study research designs. These studies demonstrate the positive effects of the approach in improving academic outcomes of learners across a variety of domains. This includes helping children become more fluent at arithmetic (see for example, Casey et al., 2003), recalling content specific terminology and definition dyads (see for example, Stockwell & Eshlelman, 2010), as well as sight reading Dolch words (Lambe, Murphy, & Kelly, 2015). More recently, comparative group studies have investigated the effectiveness of the SAFMEDS strategy against an education as usual control group (e.g., Hunter et al., 2016, Greene et al., 2018). Within these studies, a researcher with experience using the SAFMEDS strategy was present at each intervention session to support implementation and ensure high levels of fidelity.

Although sparse, there is some evidence to suggest that teachers can elicit positive student outcomes from a SAFMEDS intervention even when researchers offer no in-situ support following training. Beverley, Hughes, and Hastings (2016) acknowledged the importance of providing teachers with the training necessary for them to implement and manage a SAFMEDS intervention. Following training, the teachers participating in their study did not receive any in-situ support from a researcher to implement the strategy on a class-wide scale. The results demonstrated that the class of children who engaged with the SAFMEDS intervention made more reliable fluency progress between pre- and post-test compared to the class of children who did not use the strategy.

To date, the majority of empirical research investigating the SAFMEDS strategy focuses on efficacy designs with researcher driven implementation. Beverley et al's (2016) study suggests that teachers can elicit positive student outcomes under conditions with no researcher input following training. Whilst both of these approaches have shown positive results, it is still unclear whether researcher involvement after initial teacher training is important for implementation and children's outcomes. The aim of the current study was to provide direct insight into the impact of coaching (i.e., in-situ individualized implementation support from a researcher) during a teacher led SAFMEDS mathematics program in schools. The teachers and teaching assistants used the SAFMEDS strategy with the children in their schools to help develop fluency of arithmetic.

In our theory of action, the initial training intended to support teacher's pedagogical knowledge of the SAFMEDS strategy and the associated data-driven teaching practice (PT). We anticipated that coaching support would improve the fidelity of the teacher's implementation. A researcher tailored each in-situ visit to the individual needs of each teacher, but broadly these sessions aimed to address challenges such as: interpreting learning pictures; identifying and correcting children's procedural steps as they progressed through each SAFMEDS timing; and managing challenges such as cheating and identifying appropriate learning materials. As a result of more accurate implementation, our theory of action proposed that children attending schools where their teacher received coaching would make greater fluency progress between baseline and follow-up, compared to those attending schools that did not receive coaching following training.

In line with previous studies that have investigated the effects of the SAFMEDS strategy, the outcomes from this research relate to children's arithmetic fluency. We acknowledge that it would have been beneficial to collect data directly relating to fidelity of implementation but were unable to due to practical and funding restraints. To our knowledge, this is the first randomized controlled trial (RCT) investigating the effects of providing implementation support to teachers following SAFMEDS training. Answering this question would provide a foundation for further research investigating the mechanisms that make SAFMEDS coaching programs effective and contribute to the broader literature about the effects of coaching for teachers on intervention outcomes for students.

Method

Trial design and participants

As part of a wider initiative to improve numeracy standards across North Wales, the Regional School Effectiveness and Improvement Service for North Wales (GwE) disseminated the initial advertisement for this project. For a school to be considered eligible, they had to located within one of the six local authorities supported by GwE (Conwy, Denbighshire, Flintshire, Gwynedd, Anglesey, or Wrexham). Table 3.1 outlines the characteristics of the schools included in the randomization.

To participate schools needed to be willing to release teacher(s) to attend the training at the beginning of the project. The nominated teacher needed to be able to invest the necessary amount of time per week to deliver the SAFMEDS intervention (i.e., three 20-minute sessions). The advertisement explained that by enrolling on the project schools would be randomized to one of the two trial arms. Schools had nominated teachers to complete the training *before* they knew which trial arm they had been allocated to. Therefore, any trial arm differences in the roles of teaching staff selected for training by the schools were due to chance. Table 3.2 displays the baseline characteristics of the teachers who attended training.

Each school selected up to 10 children to participate in the SAFMEDS mathematics intervention prior to randomization. We disseminated an opt-out consent form to all the children's parents/guardians detailing the aims of the study. This form asked if we could collect and analyze their child's outcome data. In instances where the consent form was returned, teachers could still include the children in the SAFMEDS intervention, but we did not collect their anonymized data for analysis. We had consent to analyze the data from 575 children (n_{Support} = 294, $n_{No\text{Support}}$ = 281), across 60 schools (n_{Support} = 31, $n_{No\text{Support}}$ = 29).

For children in Year 3 or above (aged $\geq$ 7 years), we asked teachers to implement the intervention with children who scored less than 100 standard points on the national numeracy procedural test undertaken at the end of the preceding academic year. All children in Years 2-9 (aged 6 to 14 years) who attend a maintained school in Wales (i.e., schools funded by a local education authority) sit this formative test at the end of each academic year. Children sit the procedural test online as it offers a personalized assessment experience (i.e., the questions get easier/more challenging depending on the child's ability). The procedural numeracy test assesses all relevant aspects of the numeracy curriculum in Wales.

Children in Year 2 had not completed the national tests at the start of the study. In these instances, we asked schools to identify the children who they felt needed intervention support to improve fluency of basic mathematics skills and/or who they judged to be working below the expected standard for their age. These children were those who needed supplementary tuition to improve their fluency of arithmetic facts.

The mean age of the children attending schools randomized to the no support arm was 7-years 3-months (range: 6-years 0-months to 9-years 2-months; *SD* = 14.34 months). The mean age of children attending schools allocated to the ongoing support arm was also 7-years 3-months (range: 6-years 0-months to 15-years 10-months; *SD* = 14.32 months). It is worth noting that two secondary (high) schools participated in this study. One of these schools worked with a group of Year 7 students (aged 11-12 years) who significantly underperformed on the procedural test. The other secondary school was a special educational needs school that supported children aged 11-17 years; these children lacked basic mathematics skills (e.g., single digit addition). Table 3.3 outlines the characteristics of the children included in the randomization. Table 3.4 displays the baseline characteristics for the children's outcome measures.

Randomization

Randomization occurred after all teachers received the SAFMEDS training but prior to the children completing the baseline assessments. A statistician—who was independent to the study—randomly allocated schools to one of the two trial arms using minimization. During this allocation, the statistician stratified schools by County (local authority) and the language used predominantly for teaching (English versus Welsh medium). Some of the schools had the same headteacher; in these instances, the statistician treated the schools as one cluster to prevent bleeding effects across conditions. In terms of hierarchal structure, teachers and children were nested within each school. The first author could not be masked to the randomization due to the need to conduct support visits. However, the assessors who conducted the baseline and follow-up assessments remained blind to the allocation of each school.

Table 3.1

Characteristics of the Schools Included in the Randomization

	Variable	Ongoing support *n*	No support *n*
Stage of school	Primary	31	31
	Secondary (high)	2	-
School type	Mainstream	32	30
	Special educational needs	1	1
County/local authority	Conwy	12	14
	Denbighshire	7	3
	Flintshire	5	4
	Gwynedd	5	4
	Anglesey	1	2
	Wrexham	3	4
Predominant language	English	21	21
	Welsh	12	10

Table 3.2

Role of Staff Who Received Training at the Beginning of the Project

Job title	Ongoing support (*n*)	No support (*n*)
Head teacher	2	1
Deputy head teacher	3	3
Teacher	22	18
HLTA*	1	5
Teaching assistant	30	36
Intervention coordinator	1	-
ALNCo**	-	1

*HLTA = higher level teaching assistant **ALNCo = additional learning needs coordinator

Table 3.3

Characteristics of the Children Randomized at Baseline

		Ongoing support	No Support
	Variable	*n*	*n*
Gender	Male	162	158
	Female	152	143
Eligible for free school meals (eFSM)	Yes	75	83
	No	228	212
Predominant language	English	222	219
	Welsh	92	82
County	Conwy	128	145
	Denbighshire	48	30
	Flintshire	39	37
	Gwynedd	58	40
	Anglesey	18	10
	Wrexham	23	39
School year (age)	**2** (6-7 years)	93	163
	3 (7-8 years)	192	128
	4 (8-9 years)	4	9
	5 (9-10 years)	3	1
	6 (10-11 years)	2	-
	7 (11-12 years)	11	-
	8 (12-13 years)	1	-
	9 (13-14 years)	4	-
	10 (14-15 years)	-	-
	11 (15-16 years)	4	-

Table 3.4

Baseline Measurements for the Outcome Measures Based on All of the Children Included in the Randomization

	Ongoing support			No support		
	n	*M (SD)*	Min to max	*n*	*M (SD)*	Min to max
Measure						
MFaCTs: Grades 1-2	304	10.53 (8.88)	0 to 65	288	10.45 (8.88)	0 to 35
MFaCTs: Grades 3-5	304	7.26 (6.61)	0 to 21	288	7.19 (6.60)	0 to 26

Intervention

SAFMEDS training (all teachers)

All teachers received the same training prior to randomization. During the 3-hour training session, we introduced the teachers to some of the basic theory behind the SAFMEDS strategy, modelled the procedure (as detailed in Table 3.5), and gave them the opportunity to practice using the cards. Following four 1-minute SAFMEDS timings, we showed the teachers how to record data and graph it on a SCC. During the training, we also emphasized the importance of interpreting learning pictures in relation to children's learning progression throughout the intervention (see Lindsley, 1995). After showing the different learning pictures, we went through a series of common scenarios using SCC data from previous research projects. The scenarios prompted discussion relating to cheating, identifying skill deficits, and deciding whether something in the surrounding environment may be affecting a child's scores (e.g., missing their favorite lesson to take part in the SAFMEDS session, or a loud music lesson scheduled in the room next door). We discussed what learning pictures may develop as a result of these scenarios and suggested some interventions that might be appropriate to try (e.g., creating individualized score targets, building fluency of prerequisite skills, or changing the time/location of the SAFMEDS session).

Throughout the intervention period the children engaged with the SAFMEDS strategy via a deck of flashcards. On the front of each card was a question (e.g., 5 + 6 =) and on the back was the corresponding correct answer. During the training, we provided teachers with all the materials that they would need to start the SAFMEDS intervention in their school. This included decks of addition and subtraction SAFMEDS cards, score tables, SCCs, and a placemat (so the children could easily distinguish between their "correct" and "not yet" cards). All teachers who attended the training also had access to printable PDF materials of component arithmetic skills across the national curriculum which they could download at their convenience. We instructed all schools to focus on single digit addition skills first and then progress through card decks as required (in line with the children's learning pictures).

We instructed teachers to use the SAFMEDS strategy at least three times per week with the children they were supporting. Each session should consist of four SAFMEDS timings and last approximately 20-minutes. Within these sessions, the teachers and children had clearly defined roles. The children were to work through their cards independently during each 1-minute timing (as outlined within Table 3.5). Teachers were required to monitor aspects of fidelity (e.g., ensuring the children: followed each of the appropriate steps, were not cheating, and engaged with the sessions regularly). Additionally, we encouraged teachers to support children during the error correction step (including some one-on-one or small group teaching if necessary), review charted data regularly, and ensure that children were practicing a skill that was appropriately matched to their existing skill level. Once children had claimed they had become fluent at a deck of cards, it was also important that the teacher was able to verify this (e.g., watch a timing) before they issued a deck for a more difficult skill.

Table 3.5

An Outline of the Say-All-Fast-Minute-Every-Day-Shuffled Strategy

Timing	**Action**	**Learning principle** [corresponding action]
Before timing	1. Shuffle the cards. 2. Teacher sets a timer for 1-minute.	Prevents serial learning [1] Short focused practice sprints [2]
During timing	3. Children read the front of the card in their head and say the answer out loud. They should turn each card over to check their answer, before placing it in either their "correct" or "not yet" pile.	Active responding [3] Immediate feedback [3]
After timing	4. Once the timer has finished, the teacher says stop. 5. Children count their cards and write their scores down in the given table. 6. If a child gets any cards in their "not yet" pile, they should address these cards (error correction). 7. All cards should be put back in one pile, ready to shuffle and go again. 8. Following all four timings, the child should take their best score and plot it on their SCC. 9. At the end of each week, a teacher should look at each child's data. If they have shown little, to no, progression over three consecutive days they should consider making a change within the program.	Formative assessment [5] Practice and firm new skills [6] Repetition to build mastery [7] Assessment of learning [8,9]

Ongoing support

Schools allocated to this trial arm received three in-situ support visits from the first author throughout the duration of the study (November, February, and May). The first author had several years' experience of using the SAFMEDS strategy in schools, so was able to advise teachers on themes around implementation and interpreting the children's data. Each visit was individualized based on the needs of each teacher and the children they were supporting. Examples of support varied, but largely consisted of the following: modelling sessions; observing the teacher(s) or teaching assistant(s) delivering the intervention and providing direct feedback on implementation; suggesting interventions for children who were struggling to progress with particular decks (e.g., focusing on building fluency in prerequisite skills); discussing ways that teachers might be able to integrate the intervention more readily (e.g., adopting a peer-led approach to support error correction and reduce cheating); and supporting teachers to interpret the children's learning pictures. Each scheduled visit lasted 1-hour.

Between visits, teachers could email the first author about any issues relating to the intervention or the technology used to support the project. Teachers allocated to this trial arm contacted the author about accessing materials (18 instances; 12 schools), to gain advice about helping children progress (7 instances; 6 schools), and for advice about interpreting data (2 instances; 2 schools).

We gave all teachers—irrespective of trial arm— the option to plot the children's data using either paper or electronic SCCs. There were 10 instances (across 8 schools) where teachers allocated to receive support emailed the first author to report issues logging the data electronically. Moreover, we made all of the resources for this project available via the Welsh Government's online school platform for educational resources (HwB). Two teachers (across 2 schools) allocated to the support trial arm emailed about gaining access to the SAFMEDS HwB platform.

No support

Following training at the beginning of the project, schools allocated to this arm received no implementation support from the first author. Teachers in this condition could contact the first author if they had any technical problems accessing the resources or inputting data into the electronic charts; but were not able to seek advice regarding the day-to-day implementation of the SAFMEDS strategy. There were 9 instances (across 7 schools) where teachers emailed the author to request access to resources, 2 instances (across 2 schools) where teachers needed support accessing the SAFMEDS HwB platform, and 10 instances (across 9 schools) where teachers reported issues logging their children's data electronically.

During the training, we highlighted an additional caveat about the support we could offer schools allocated to this arm. We had an ethical obligation to provide the teachers with support if they felt like they could not initiate or sustain the intervention without it. No school in this condition asked for additional support, but if they did, we would have provided it and handled their data appropriately. It is also important to note that the "no support" group was essentially a "support as usual" group. Typically, schools would seek a training course, send their staff on the course, and then implement interventions on their own (unless they specifically purchase additional support with implementation). Thus, we believe that the no support trial arm is an ecologically valid comparison for inclusion within this study.

Procedure

Baseline assessments

The children completed the Mathematics Fluency and Calculation Tests (MFaCTs; Reynolds, Voress, & Kamphaus, 2015). The Grades 1-2 fluency assessment measures addition and subtraction fluency and is intended for children aged between 6-years 0-months and 8-years 11-months. The Grade 3-5 fluency assessment measures addition, subtraction, multiplication, and division fluency; this assessment is intended for children aged between 8-years 0-months and 11-years 11-months. We used both measures with all the children in the sample to provide an inclusive overview of their skill progress across the intervention. To reduce practice effects, the MFaCTs assessments offer parallel test forms. The published statistics for these tests show high internal reliability across ages ($\alpha > .80$). We used Form A during the baseline assessments.

The children came out in a group to complete these assessments but filled in their forms individually and in silence. The assessors provided each child with a pencil and the test form. The children completed MFaCTs: Grades 1-2 first. They had 5-minutes to answer as many of the 100 questions on the page as they could; working across the page from left to right. If they did not know the answer to a question, they were allowed to skip it and move onto the next one. Once the timer finished, an assessor instructed the children to turn the form over so that they could collect them. The children then repeated this procedure for the MFaCTs: Grades 3-5 assessment.

Eight-month follow-up assessments

Eight months post-randomization, we reassessed the children who participated in the study. This process mirrored the administration of the baseline assessments, with the children completing both MFaCTs fluency assessments (Form B).

Figure 3.1 outlines the flow of participants from enrolment to the final analysis. Prior to the follow-up assessments, four schools indicated that they were no longer using the SAFMEDS intervention due to unforeseen challenges with staffing. Three of these schools were happy for us to still collect follow-up data from their children (denoted as intend to treat); whilst one school was unable to accommodate this (denoted as withdrawal).

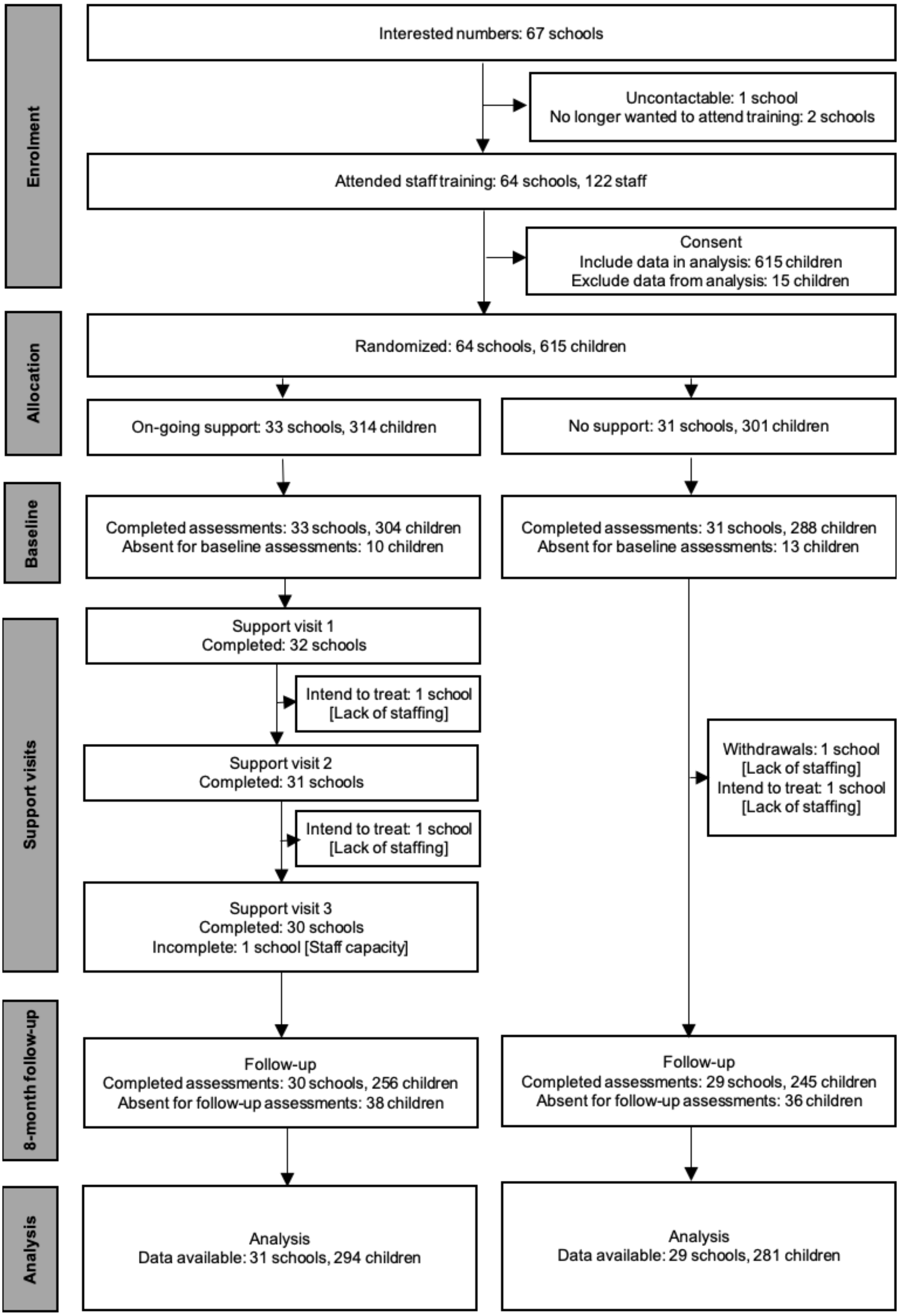

Figure 3.1. CONSORT flow diagram

Analysis

The data for this study falls within two hierarchal levels (level 1 = children, level 2 = school). Due to children being nested within schools, we analyzed the data using a multi-linear mixed effects model. This analysis is consistent with other studies that have adopted cluster RCT designs (see, for example, Jahoda et al., 2017; Zimmermann et al., 2014). Linear mixed-effect models enable analysis of continuous outcome variables within hierarchal research designs by partitioning the overall variance of the outcome variable into factors that correspond to the different levels of the hierarchy (Galecki & Burzykowski, 2013). Baayen, Davidson, and Bates (2008) further outlined some of the advantages of using mixed effects modelling over univariate alternatives, such as ANOVA or ordinary least squares regression.

Due to lack of availability of standardized scores for the range of ages included within this sample, we opted to analyze the children's raw scores on the MFaCTs measures. We used Stata v13.0 to analyze the raw data from this trial. Using Xtmixed, we assessed the interaction between time (baseline versus follow-up) and trial arm (ongoing support versus no support) across the fluency (MFaCTs) measures. Level 1 within our model contains covariates associated with individual children, these were: gender, predominant home language, eligibility for free school meals status (eFSM), and school year group. Level 2 within our model refers to the covariates associated with each school, these were: school administrative county, trial arm, and time. The model also generated the intraclass correlation coefficients values (ICCs) associated with each level of the model.

To assess the impact of support, we calculated a Cohen's *d* effect size for each measure. To calculate Cohen's *d* and the associated 95% confidence intervals we adhered to Feingold's (2015) formulae. We have discussed the outcomes of the results in relation to Cohen's (1988) benchmarks, whereby an effect is small ($d = 0.20$ to $d = 0.49$), medium ($d = 0.50$ to $d = 0.79$), or large ($d \geq 0.80$).

We carried out sensitivity analyses by repeating the main analysis using multiple imputation and a complete cases analysis approach. The effect sizes varied minimally (refer to supplementary material). Existing published data suggest that certain factors predict an attainment gap between sub-groups of school-aged children. These include differences in outcome variables across genders and levels of social deprivation (OECD, 2012). Moreover, Van Rinsveld, Dricot, Guillaume, Rossion, and Schiltz (2017) provided evidence to suggest that bilingual individuals rely on differential activation patterns in the brain to solve simple and complex arithmetic questions in their different languages. As such, we also conducted a series of moderation analyses to investigate the effects of these variables (refer to supplementary material). We found no evidence of these factors moderating the effect of trial arm on children's mathematics outcomes.

Results

Support model

Figure 3.1 outlines the number of schools who completed each support visit. By the final visit, two schools allocated to receive ongoing support had stopped using the SAFMEDS strategy due to unforeseen changes to staff availability. Of the schools continuing to use the SAFMEDS strategy, all but one engaged with the three support visits. Seventeen schools allocated to the support arm made email contact with the first author between visits to access further support.

Fluency outcomes

We were interested in investigating whether implementation support from a researcher could help improve children's fluency outcomes during a teacher led SAFMEDS intervention. In terms of the MFaCTs: Grades 1-2 assessment, the statistical analysis revealed a small positive effect of ongoing support over no support on the children's addition and subtraction fluency between baseline and follow-up (Trial arm x Time: $\beta = 2.92$, $SE = 0.86$, p

= .001, $d = 0.23$). A pairwise comparison of marginal linear predictions, with Bonferroni correction, revealed significant improvements on this measure for children in both arms. Children's raw scores in the support arm improved to a greater extent on average between baseline ($M = 12.00$) and follow-up ($M = 22.59$; $p < .001$) compared to children in the no support arm ($M_{\text{baseline}} = 9.02$, $M_{\text{follow-up}} = 16.50$, $p < .001$).

Analysis of the MFaCTs: Grades 3-5 showed that ongoing support has a small positive effect, relative to no support, on the children's addition, subtraction, multiplication, and division fluency ($\beta = 2.68$, $SE = 0.75$, $p < .001$, $d = 0.25$). Bonferroni-corrected pairwise comparisons revealed that children's raw scores on this measure improved significantly in both arm of the study. Children in the ongoing support arm improved to a greater extent between baseline and follow-up ($M_{\text{baseline}} = 8.52$, $M_{\text{follow-up}} = 19.12$, $p < .001$) than children in the no support arm ($M_{\text{baseline}} = 5.94$, $M_{\text{follow-up}} = 13.76$, $p < .001$). *Table 3.6* displays further descriptive statistics from the linear mixed effects analysis for both MFaCTs measures.

Table 3.6

Summary of the Model Outcomes

	Ongoing support		No support		Adjusted cases analysis*				
	n	Marginal mean of raw scores [95% CI]	*n*	Marginal mean of raw scores [95% CI]	Adjusted difference between change scores [95% *CI*]	*p*	*d* [95% CI]		ICC
MFaCTs: Grades 1-2									
Baseline	285	12.00 [10.28, 13.73]	269	9.02 [7.28, 10.76]	-	-	-	School	0.36
Follow-up	256	22.59 [20.85, 24.34]	237	16.50 [14.72, 18.29]	2.92 [1.22, 4.61]	.001	0.23 [0.06, 0.39]	Children	0.46
MFaCTs: Grades 3-5									
Baseline	285	8.52 [7.00, 10.05]	269	5.94 [4.40, 7.48]	-	-	-	School	0.37
Follow-up	251	19.14 [17.59, 20.69]	238	13.76 [12.19, 15.34]	2.68 [1.21, 4.14]	<.001	0.25 [0.08, 0.43]	Children	0.42

* Model adjusted for the following fixed effects: county, child's predominant language, gender, free school meals status, school year group. Interaction effect: trial arm x time. Random effects: school, children.

Discussion

Our aim for the current study was to gain insight into the putative benefits of providing teachers with implementation support throughout a SAFMEDS mathematics program. An increasing number of teachers across North Wales are using the SAFMEDS strategy to support children's fluency of basic mathematics skills. Yet, no known research internationally had investigated whether implementation support from a researcher can lead to better fluency outcomes than the more traditional "no support" approach following teacher training. Identification of a successful coaching model could help researchers to support this program at scale, help teachers advance their PD, and improve the outcomes of the children they teach. The results from this cluster RCT suggest that providing teachers with initial training in SAFMEDS and then three 1-hour visits and email contact with a researcher has a positive effect on children's fluency of arithmetic facts compared to initial training only. This paper also contributes to the growing literature reporting the effects of coaching teachers to implement evidence-based interventions within their schools, with a specific focus on mathematics outcomes.

The Education Endowment Foundation (EEF, 2019) are a leading UK charity that support the generation of research and good practice within schools. They aim to support teachers to use evidence that works to improve educational outcomes for children. In their recent implementation guidance report, the EEF identified the importance of reinforcing initial training for interventions with expert follow-on support within school. The results from our study further support this guidance in the context of a SAFMEDS mathematics intervention. Whilst children attending schools in the no support arm of this trial did improve their fluency of arithmetic facts, children made more significant progress when their teacher received coaching to support their implementation of the SAFMEDS strategy.

Data from Kraft et al. (2018) supported the idea that someone with expertise can coach teachers to implement evidence-based interventions in schools. The results from their meta-analysis revealed that PD programs with an element of coaching can have a positive effect on student achievement outcomes by +0.18 *SD*; although this largely reflected their application to literacy and content-based interventions. In contrast, Lynch et al's (2019) meta-analysis suggested that there was no added benefit to having a coaching element as a part of the format for PD interventions for mathematics and science. In the current study, we provided direct experimental manipulation of a coaching element to the SAFMEDS intervention and found the effect of coaching to be between +0.21 and 0.23 *SD* across the MFaCTs measures; these outcomes are similar to Kraft et al's findings. Our results provide some additional support for the effectiveness of teacher coaching in the context of a fluency-based arithmetic intervention.

Kraft et al's (2018) analysis also revealed that effect sizes varied significantly depending on whether the researchers devised their own assessments or administered standardized tests. When considering effect sizes within education research, Kraft (2020) outlined that researcher-designed assessments often reflect content that more closely align with the outcomes of the evaluated program, compared to the broader scope of standardized assessments. The MFaCTs measures are published and standardized, however the focus on fluency of arithmetic facts aligned closely with the content the children covered within the SAFMEDS sessions. This may have inflated the observed effect sizes compared to alternative standardized assessments.

When interpreting the results from the current study, it is important to consider the underlying mechanisms and social contingencies that might have made coaching effective. First, the support visits served to provide teachers with feedback to improve their implementation fidelity. Durlak and DuPre (2008) reported that without support teachers are often unable to implement an intervention to 100% fidelity following training. This is not surprising given that field studies come with additional extraneous variables compared to efficacy/laboratory designs (Cook & Odom, 2013). However, improved implementation fidelity of evidence-based interventions in the classroom can lead to improved student outcomes (Durlak & DuPre, 2008; Ysseldyke et al., 2003). By providing teachers with in-situ support during the present RCT, our aim was to help them deliver the program in a way that more closely aligns with its intended design (e.g., ensuring the children engaged with the practice regularly, discussing methods to address and reduce cheating, as well as reviewing and acting upon children's progress data). Improved adherence to the procedural aspects of the program may explain why the children who attended schools allocated to the support trial arm made greater fluency gains.

Second, between each visit, the teachers had the opportunity to adapt their practice based on the feedback they had received. By design, PT practices allow teachers to monitor and reflect upon their children's learning (Lindsley, 1995). If children are not making desirable progress towards fluency, then their teacher *should* adapt the instruction or materials that they provide. Through session observation and review of these data, the first author would have been able to see progress across the program. As such, there is a level of accountability that the teachers might have experienced to avoid feeling embarrassed during the following support visit. In a qualitative evaluation of a coaching program in a healthcare setting, Liddy, Johnston, Irving, Nash, and Ward (2015) reported that coaches helped patients realize that they need to play an active role in managing and improving their health. The patients also reported that their personal accountability increased as a result of their engagement with the coaches because they knew someone else was monitoring their engagement with the program. It seems feasible that this finding could extend to programs relating to school-based educational interventions. It is difficult to disentangle implementation fidelity and accountability, but both of these mechanisms provide direction for future research in this area.

We acknowledge that this study would have been enhanced if we collected data relating to teachers' and children's implementation fidelity across both trial arms. By employing blind observers to attend a SAFMEDS session in each school following each cycle of support visits, it would have been possible to directly assess the effects of expertise on the teacher's implementation fidelity. Moreover, analysis of these data could identify common aspects of the strategy that teachers struggle, or fail, to implement in school settings. Due to practical constraints and funding, we were unable to incorporate this into the current study. However, this would provide a valuable extension to future replications. It is possible that the implementation support offered by the researcher helped teachers to: interpret learning pictures more readily and accurately; identify and correct children's procedural steps as they progressed through each SAFMEDS timing; as well as manage challenges such as cheating and identifying appropriate curriculum materials.

Whilst we did not carry out a formal economic analysis, we believe that this support model may be a cost-effective and feasible alternative to embedding a researcher in each school to run and maintain a SAFMEDS intervention. Adoption of the current support model would enable a researcher to provide necessary implementation support at scale and may encourage teachers to use the intervention beyond the termination of a research study. Costs associated with the replication of this support model include a researcher's time (three 1-hour support visits and designated time to respond to email queries), cost of travel between schools, the cost of materials (e.g., printing each SAFMEDS deck double-sided onto card; at approximately 6 sheets of A4 card per deck per child), and the cost a teacher/TAs time to prepare and deliver three SAFMEDS sessions per week (with each session lasting approximately 20-minutes).

The results from the current study suggest that initial training can provide teachers with skills to implement the SAFMEDS strategy in their school. Children across both trial arms evidenced improvements in their fluency of arithmetic facts between baseline and follow-up across both MFaCTs measures. Support from a researcher helped teachers to elicit greater fluency progress from the children that they worked with. Further research is still needed to establish the components of this model that make the support effective; including the exploration of the effects of coaching on teacher’s implementation fidelity and perceived accountability.

Chapter 4 : Assessing the social validity of the SAFMEDS strategy from the perspective of teachers and children

Owen, K., Watkins, R., Beverley, M., & Hughes, J. C (in prep). *Assessing the social validity of the SAFMEDS strategy from the perspective of teachers and children.* Manuscript in preparation.

Preface

After spending time interacting with teachers and children during the support visits of the c-RCT (Chapter 3), the first author was interested in exploring the social validity that these stakeholders associated with the SAFMEDS strategy. When thinking about implementation science and *why* teachers may not to continue using the strategy within their school beyond the termination of the research strategy, the first author was interested in exploring some of the challenge's teachers may face with implementation. She was also interested in documenting the experiences and opinions of the children who have use the strategy to support their fluency of arithmetic. To date, there are no known published papers that report a robust evaluation of the social validity associated with using the SAFMEDS strategy in schools. Figure 2.3 provides a graphic representation of the context, aims, and process underlying this study.

This chapter contains the draft manuscript of the paper that we intend to submit for publication. We have edited the formatting in line with APA guidelines (including placing figures and tables in the main body of the text). In Chapter 7 we discuss the implications, applications, and limitations of this research in more detail.

Abstract

The Say-All-Fast-Minute-Every-Day-Shuffled (SAFMEDS) strategy promotes fluency across a number of skills and contexts. However, few studies have reported the social validity key stakeholders associate with using the strategy in schools. Study 1 details the findings from a survey completed by teachers (*N* = 55). Using thematic analysis, we identified five themes: (1) factors that promote progress; (2) factors that limit progress; (3) impact of competition; (4) confidence; and (5) inherent advantages of the SAFMEDS strategy. These themes provide insight into teachers experiences of implementing the strategy under the real-word conditions of the classroom and a range of accompanying advantages and potential challenges they may face. Within study 2, we discuss themes arising from interviews with children (*N* = 26) about their views and experiences using the SAFMEDS strategy. These children had used the strategy with their teacher for one academic year to promote fast and accurate recall of arithmetic facts. Analysis of these transcripts revealed five further themes relating to children's engagement with the strategy: (1) enjoyment; (2) data; (3) sense of achievement; (4) skills; and (5) home use.

Say-All-Fast-Minute-Every-Day-Shuffled (SAFMEDS) is a practice and assessment strategy derived from Precision Teaching (PT; Potts, Eshleman, & Cooper, 1993). PT is a system for defining and measuring fluency of skills on a child-by-child basis (Kubina & Yurich, 2012). By adopting the PT approach, educators make data-based decisions about subsequent instruction through the analysis of learning pictures that emerge on a child's Standard Celeration Chart (SCC; Binder & Watkins, 1990; White 1986). Lindsley (1995) outlined 13 learning pictures that can emerge from data on an SCC, which can be further categorized into three types of progress pictures: improving, maintaining, and worsening. If upon visual inspection, a child had made little or no progress towards their fluency aim (i.e., their data depicts a maintaining or worsening picture) then their teacher may need to alter the instruction or materials that they provide. Adhering to the PT approach allows educators to make real-time decisions and prevent prolonged periods of non-progression (Merbitz, Vieitez, Merbitz, & Pennypacker, 2004).

Learners typically practice the SAFMEDS strategy using a deck of flashcards; with a stimulus on the front (e.g., a mathematics question) and the corresponding correct response on the back (Meindl, Ivy, Miller, Neef, & Williamson, 2013). The learner reads the front of the card silently before vocalizing the answer (Quigley, Peterson, Frieder, & Peck, 2017). They then receive immediate feedback by checking their answer with the back of the card (Lindsley, 1996b). The aim is to get through as many of the cards as quickly as possible whilst separating the correct responses from the errors (Cihon, Strutz, & Eshleman, 2012). Previous research has demonstrated that, when paired with data-driven decisions about instruction, the SAFMEDS strategy can improve skill fluency across several academic domains. For example, learners can use the SAFMEDS strategy to improve accuracy and recall of arithmetic facts (Casey, McLaughlin, Weber, & Everson, 2003; Cunningham et al., 2012; Hunter et al., 2016; Nam & Spruill, 2005), words in a second language (Beverley, Hughes, & Hastings, 2016; Bolich & Sweeney, 1996), and subject-specific terminology (Beverley, Hughes, & Hastings, 2009; Stockwell & Eshelman, 2010; Meindl et al., 2013).

The existing literature suggests that the SAFMEDS strategy is suitable for use in schools. Practitioners have used the strategy to produce impactful academic outcomes on a variety of scales from one-on-one practice (Cunningham, McLaughlin, & Weber, 2012) to class-wide (Hunter, Beverley, Parkinson, & Hughes, 2016); with both mainstream learners (see, for example, Greene, Mc Tiernan and Holloway, 2018) and with children who have additional learning needs (see, for example, Casey et al., 2003). Following training, teachers can implement the strategy in their classrooms to improve academic outcomes (Beverley et al., 2016; Owen et al., under review). Together, the preceding research suggests that the SAFMEDS strategy can be successfully adapted for use across a range of school contexts and classroom needs.

Despite a developing quantitative evidence-base supporting the use of the SAFMEDS strategy in schools, no known studies have robustly evaluated the social validity key stakeholders associate with using it. Assessing social validity enables us to evaluate the effectiveness of an intervention/program beyond the objectively measured target behaviours (Baer, Wolf, & Risley, 1987). The available SAFMEDS literature largely focuses on children's fluency and academic development across an intervention period. Hunter et al. (2016) asked students who had used the SAFMEDS strategy to raise their hands if they would like to continue using it. Approximately 90% of the students in their sample (n = 19) indicated that they would like to continue using the strategy. However, Hunter et al. identified that a more robust evaluation of student views is necessary to establish the social validity of the SAFMEDS strategy. No known research has reported teachers views on the strategy.

This paper details two qualitative studies focusing on the social validity of the SAFMEDS strategy from the perspective of teachers (study 1) and children (study 2) who have used it in schools across North Wales. Specifically, the questions that we asked aimed to gain insight into teacher's experiences of implementation, perceived usefulness of the strategy, and factors that may affect children's engagement.

Study 1: Teachers' views and experiences

Method

Ethics

This study received ethical approval from Bangor University's ethics committee (application number: 2018-16309). The first page of the survey outlined the aims of the research and details about what would happen to respondents' data. This page stated that the research team were interested in collecting data relating to teachers' and teaching assistants' (TAs) views and experiences of using the SAFMEDS strategy within their schools. All participating members of teaching staff had to provide consent before beginning the survey. They had the option to leave questions blank if they did not wish to answer them. Throughout this paper we have not referred to the teaching staff or schools by name to protect their anonymity.

Recruitment

The first author emailed all members of teaching staff who had attended a SAFMEDS training session between the years 2016 and 2018 with a link to the online survey. These training sessions had been organized by the Regional School Effectiveness and Improvement Service for North Wales (GwE) in collaboration with researchers at Bangor University. As such, all staff received training from a researcher to follow the SAFMEDS strategy in line with the procedural steps outlined in Table 4.1.

Table 4.1

An Outline of the SAFMEDS Strategy (Table Adapted from Chapter 3)

Timing	Action
Before timing	1. Children shuffle the cards. 2. A member of staff sets a timer for 1-minute.
During timing	3. Children read the front of the card in their head and say the answer out loud. They should turn each card over to check their answer, before placing it in either their "correct" or "not yet" pile.
After timing	4. Once the timer has finished, the member of staff says stop. 5. Children count their cards and write their scores down in the given table. 6. If a child gets any cards in their "not yet" pile, they should address these cards (error correction). 7. All cards should be put back in one pile, ready to shuffle and go again. 8. Following all four timings, the child should take their best score and plot it on their standard celeration chart (SCC). 9. At the end of each week, a member of staff should look at each child's data. If they have shown little, to no, progression over three consecutive sessions they should consider making a change to the instruction/task.

Sample

Fifty-five members of teaching staff from schools within North Wales completed the online survey. Refer to Table 4.2 for a list of their roles within the school. Table 4.3 outlines the skills that the children in their schools were practicing and assessing using the SAFMEDS strategy.

Table 4.2

Teaching Roles of the Staff Who Completed the Online Survey

Role	*n*
Teacher	12
Teaching assistant (TA)	26
Higher Level Teaching Assistant (HLTA)	5
Deputy headteacher	8
Headteacher	3
Unqualified teacher	1

Table 4.3

Skills the Staff Supported Using the SAFMEDS Strategy

Skill(s)	Count
Numeracy	52
Numeracy and literacy	2
Unspecified	1

Procedure

Teachers had the option to complete the survey through the medium of English (*n* = 46) or Welsh (*n* = 9) at their own convenience. Five questions collected demographic information (i.e., name, school, job title of the respondent, role of the staff member who ran the SAFMEDS sessions, and which skills they targeted using the SAFMEDS strategy). The remaining seven questions aimed to gain insight into the advantages and challenges associated with implementing a SAFMEDS program in schools (for a full list of questions, refer to Appendix C). We used open-ended questions so that each respondent could to provide as much or as little detail as they wanted to for each question. The survey also contained space for staff to justify whether they would use the strategy in the future and to report any specific case studies that they wanted to share. We translated the Welsh survey responses into English prior to analysis.

Analysis

The first author adhered to the six stages of thematic analysis outlined by Braun and Clarke (2006). Due to collecting the survey responses electronically, there was no need to manually transcribe the data. In order to become familiar with the dataset the first author printed out each complete survey and read through each response twice before beginning analysis. On consecutive readings the first author made a note of any salient and interesting comments within the dataset, along with some initial thoughts about emerging codes. By working through each response within the transcript in detail, the first author was able to capture some of the key codes across the data (for an example of this process, refer to Table 4.4).

Table 4.4

An Example Data Extract from the Online Survey and the Corresponding Codes

Data extract	Coded for
Working in small groups enables the pupils to concentrate better. Also, with the answers being on the back of the cards, immediate feedback as to whether pupils have got the answer correct is beneficial for their learning. Also, the language used in SAFMEDS is more positive, for example using 'not yet' instead of wrong or incorrect.	• Group size • Concentration • Immediate feedback • Terminology

The first author transferred each code onto a post-it note to allow her to begin identifying candidate themes amongst the survey responses. After refining the themes and cross checking that they encapsulated the original data extracts, the first author named each theme and built the narrative presented within the result section of this report. Figure 4.1 represents the final themes and their relationship to one and other.

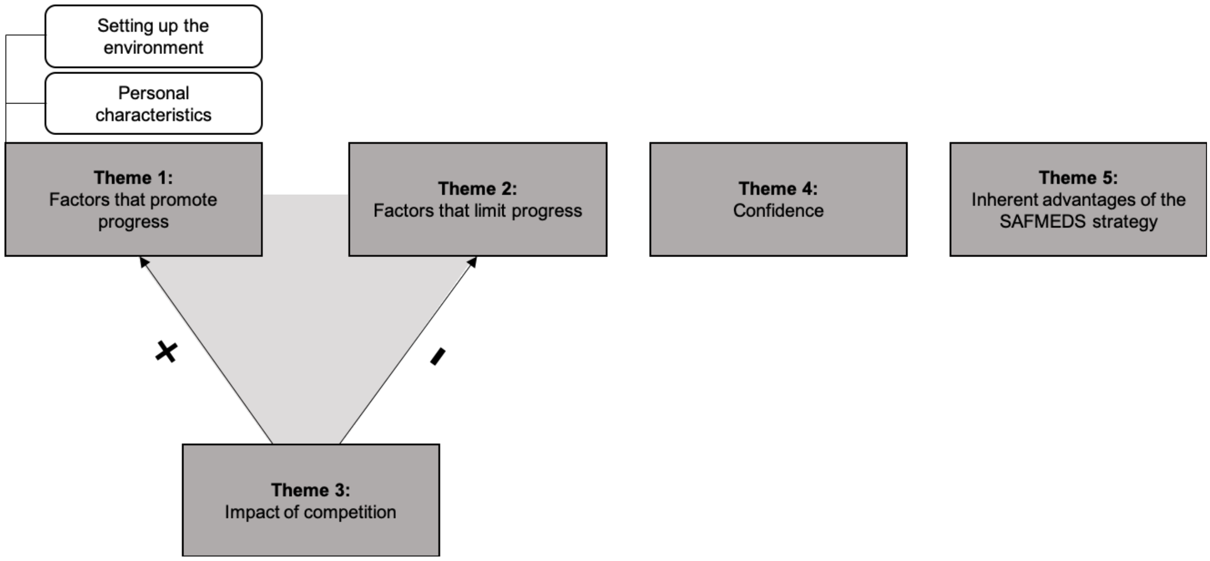

Figure 4.1. Thematic map of the themes within the online survey dataset

Results

Theme 1: Factors that promote progress

Setting up the environment

The teaching staff explained that setting up a SAFMEDS program in schools requires a transitionary period; whereby the children need to get used to using the cards and following the prescribed stages of the strategy. Following this phase, the children are typically able to take ownership of their own learning and participate in the sessions with minimal input from an adult.

> Good structure and programme runs itself once the children are aware of the routine. [Deputy headteacher]
>
> A big factor of the success is giving the intervention time. Once the children get into the routine they can do the session with minimal adult support. [HLTA]
>
> It was difficult at the beginning for them to use the cards – they would drop them or mix them up with the person sitting next door to them. Now this is not an issue as they are used to dealing with the cards. [Deputy headteacher]

The teaching staff also highlighted the importance of embedding the program into the day-to-day timetable of the school. Holding the SAFMEDS sessions on the same days of the week, at the same time, and in the same place helped staff to ensure that the children could regularly engage with the strategy. Without this routine in place, staff reported that they would miss sessions, other events would take priority, and the children would make limited progress.

> Sessions held at the same time, and in the same place by the same person each day. Quiet room – pupils were responsible for recording their own results. [Deputy Headteacher]
>
> As the sessions became regular the children increased in confidence and started to increase the number of correct responses. One of them is nearly hitting 50 correct responses per minute and is now ready to move onto the next stage. [HLTA]
>
> Pupils sometimes miss out on SAFMEDS that day which hinders the speed they progress. [TA]

When considering the logistics of a SAFMEDS program, several members of teaching staff commented on advantages of working with smaller group sizes. Teachers reported a preference for running the intervention on a one-one-one or small group basis to ensure that the children receive appropriate attention and support. This was a particularly common view of teaching staff who reported targeting children with additional learning needs. Five teaching staff also indicated that the intervention could be scaled-up to work with larger groups once the children are able to follow the prescribed steps independently.

> We would consider using SAFMEDS in the future, but in a different way to how it's used now. I'd like to see an example of how SAFMEDS is implemented across a year group, delivered as a whole teaching session as opposed to 1-2-1. [Teacher]
>
> Yes, I would [use SAFMEDS again], but I would like to use it in smaller groups as I find that in a whole class environment the pupils get frustrated with the waiting for others. [Teacher]

In order to achieve the best results from the program, teaching staff reported that it was important to have an appropriate environment to run the SAFMEDS sessions. This included having a designated quiet space for the children to work in. In schools where space was limited, staff found a reduction in children's ability to concentrate due to surrounding events (e.g., other children walking between classes).

> The regular number practice in a quiet environment helped the success of the SAFMEDS session. [Teacher]

> Distractions [can hinder the success of a SAFMEDS session] as we work in the corridor, children walk past and often classes go out which can be noisy. [TA]

Personal characteristics

Children's attitudes towards each session played a role in the overall success of the SAFMEDS program. High levels of concentration from the children helped them to engage with the timings and improve their scores. Many of the teaching staff highlighted that the children perceived the sessions to be enjoyable; which in turn motivated the children to attend and engage with the strategy.

> The pupils have got used to being organized, on time and worked well. They have been focused and tried their best. The majority have been better focused in class. [HTLA]

> The children have enjoyed attending the sessions, and we have seen a marked improvement in their confidence and ability. [Deputy headteacher]

> The pupils are thoroughly enjoying their SAFMEDS sessions, they are learning while having fun. Their ability to answer multiplication facts quickly without having to calculate the answer is definitely improving already. [Deputy headteacher]

> The children have responded positively to the sessions and are visibly enthusiastic about attending. [TA]

Theme 2: Factors that limit progress

Whilst several factors impact the success of a SAFMEDS intervention, staff also highlighted several factors that can hinder performance gains. Many of the schools ran the intervention with smaller groups of children, which often meant withdrawing them from class to complete the SAFMEDS sessions. Staff indicated that some children would refuse to attend if there was a competing, more appealing, activity available in the classroom. Moreover, other factors such as staff availability and the children being absent from school affected attendance to SAFMEDS sessions.

> [Children will not come out for the SAFMEDS session because there is a] fun activity for rest of class OR worry they won't finish their work in class or miss out on learning if they come out: both cause resistance and annoyance towards SAFMEDS. [TA]

> Other events happening in the school so that the room was unavailable or TA having to cover another TA in class. Pupil absence could be a factor [that hinders progress] too. [Deputy headteacher]

Due to the self-directed nature of the SAFMEDS strategy, staff noted that some children cheated during the sessions. Staff provided the example of children placing cards in their "correct" pile that they had either answered incorrectly or not at all. In a group scenario, one TA noted that the children engaged in cheating in an attempt to avoid being teased by their peers.

> Keeping an eye on children that place cards in their yes box they may not have necessarily known the answer! This was better when mentors started to be involved. [TA]

> Cheating took place at the start that has been nipped in the bud. At the start there was the "I scored better than you situation" where the children teased anybody that scored low again that has been nipped in the bud. [TA]

> Sometimes if the child isn't as confident as another or not as quick it can put them off or "cheat" so they can put extra cards in the yes pile. [TA]

Theme 3: Impact of competition

The impact of competition was a salient theme that emerged from the transcripts, which the authors further sub-divided into the positive and negative effects. This theme is interrelated to both promoting and limiting children's progress within SAFMEDS sessions.

Positive

Self-competition resulted in children setting individualized targets and monitoring their scores across sessions. Staff reported this as a good method of promoting healthy competition; which seemingly helped keep the children engaged and stay motivated. Some members of staff reported that some healthy peer competition helped children to improve their scores across sessions.

> It is also good for them to see the progress they have been making throughout the session and also throughout the number of weeks we have been running SAFMEDS. This gives them an incentive to want to improve and get more answers correct next time. [Teacher]

> A couple of my group have low self esteem and I think that they have gained confidence when they see their results improving. There is a healthy competitive spirit between a couple of them. [HLTA]

> [Child name] who is very nervous and had no confidence didn't like it at the start, but as the sessions progressed, he made a good leap in his scores and now dances around when he surpasses his last best score. It is cute to watch him counting how many he got right and then to see the delight on his face when he realizes he beat his last best score. [TA]

Negative

Peer competition within groups fostered a negative attitude towards the SAFMEDS sessions for some children. In some cases, this led to low-level bullying behavior towards children achieving low scores. Some of the survey responses also alluded to feelings of inadequacy if a child did not receive the highest score within the group. Peer competition also appeared to influence the children's decision to cheat in order to write down a higher score.

> As in any group children become obsessed with doing better than other children in the group. Making fun of children who under perform. [TA]

> The competitive nature of the programme can also have a negative impact on the pupils as they can get upset if another pupil achieves a higher score. [Deputy headteacher]

> If two characters have had a fallout; competing over who has the better score when they were all on the same pack put some children down; [the children started] turning cards slightly early to beat their score. [TA]

Theme 4: Confidence

The survey responses revealed that children's confidence improves throughout a SAFMEDS program; particularly with regards to the skill(s) that they had been practicing. As a result, children become less anxious about answering cards incorrectly and feel confident enough to use their knowledge outside of SAFMEDS sessions. Staff indicated that they had observed the children exchanging their new knowledge with teachers and being more willing to answer questions in class.

> Pupils enjoy going to SAFMEDS and have become more confident in their own ability. They are not worried about getting the answer incorrect and will all have a go most of the time. [Teacher]
>
> The children have enjoyed attending the sessions, and we have seen a marked improvement in their confidence and ability. [Deputy headteacher]
>
> I think their confidence increased in that subject, but also increased their independence and I often see the children passing on their SAFMEDS knowledge to other children. Before SAFMEDS I don't think any of the children who took part would have willingly guided other children. [HLTA]
>
> [The children are] more confident in class maths lessons - more willing to put hand up in class. [TA]

Theme 5: Inherent advantages of the SAFMEDS strategy

It is easy for children to learn how to follow the steps of the SAFMEDS strategy. Additionally, each timing takes 1-minute to complete which makes the procedure quick to implement. Once in a routine, staff found the procedure to be convenient to use with the children they were supporting.

> Speed and convenience make this an easy resource to use. [Headteacher]
>
> It is a short burst of intervention, its quick paced, you can see the instant response. The children revisit the same skill lots of times. [HLTA]

Developing fluency is the ultimate goal of a SAFMEDS program. As such, staff reported a noticeable difference in the speed and accuracy in which children could perform skills. In schools where the target was to improve fluency of basic arithmetic facts, staff reported that the children's mental recall of these facts improved across the intervention.

> Overall, I have seen a marked difference in the children's confidence and fluency with single digit addition. [Teacher]
>
> Improved the speed and accuracy of addition and subtraction in some children. [TA]
>
> The children have improved in their number recall and number facts which has been noted on by their teachers. [Teacher]

As a result of improved fluency in prerequisite skills, children were able to progress onto more advanced skill (e.g., move from practicing number recognition to single digit addition sums). Staff also reported that the children developed other complementary skills, such as problem solving. Due to the inherent advantages of the SAFMEDS strategy, all but three members of staff who completed the survey indicated that they would like to use the strategy in future. Two members of staff did not respond to this question, and the remaining teacher stated that the intervention was too time consuming to align with their school schedule.

> Pupil confidence has improved in terms of number recognition and number bonds. They are able to apply these in lessons in class and parents have reported that pupils are talking about maths in a more positive way. [Deputy headteacher]
>
> Pupils are more willing to respond in mental mathematics lessons. Pupils' problem solving work improved as a result. [Teacher]
>
> [Child's name] is now able to recognize most numbers independently. Her confidence with single digit is also growing. [Teacher]
>
> [A child has] gone from +1 and +2 to now having gone through the whole single digit addition pack is working on the subtraction pack. [TA]

Discussion

Previous research has demonstrated that the SAFMEDS strategy can be used in schools to promote fluency across several subjects (e.g., Beverley et al., 2016; Casey et al., 2003). Yet, no known research has investigated the social validity of the strategy from the perspective of teachers. This study aimed provide insight into this gap in the SAFMEDS literature; with a specific focus on teacher led implementation of the strategy. Through analysis of survey responses, we were able to identify five themes relating to the advantages and challenges of using the strategy in schools.

Staff identified that setting up a SAFMEDS program requires a transitionary period. During this period, children need to familiarize themselves with how to handle the cards and staff need to establish a routine for the SAFMEDS sessions. This may pose a challenge in terms of the teacher led maintenance of a SAFMEDS intervention within the classroom. Westfall, Mold, and Fagnan (2007) suggested that some manualized interventions do not align with the day-to-day challenges within real-world settings. Problems arise when experimental control is reduced, and confounding variables can more readily influence an intervention. In the context of a SAFMEDS program, teaching staff might find it difficult to establish a routine if confounding variables (e.g., competing internal school events) prevent sessions from occurring regularly. It is also important to consider the fidelity of implementation during a teacher led SAFMEDS program after the initial transitionary period. Glasgow, Lichenstein, and Marcus (2003) identified that under real-world conditions stakeholders often modify intervention protocols. Lack of adherence to the manualized steps of an evidence-based intervention programs often results in smaller effect sizes relating to academic outcomes (Durlak & DuPre, 2008). Future research should consider how we can best support teaching staff during the initiation and maintenance of a SAFMEDS program;

particularly with regards to helping staff maintain high levels of fidelity whilst establishing the initial routine.

Despite some initial challenges setting up the intervention, teaching staff identified that children developed independence as the weeks progressed. Consequently, the children were able to participate in the intervention with minimal adult supervision. Graf and Auman (2005) promoted the idea that children should take ownership of their own learning during a SAFMEDS program. Whenever possible, children should hold and direct their own cards and be responsible for recording their own data. Encouraging children to take ownership of their learning also fosters autonomy as they have a choice as to whether they engage or not (Deci & Ryan, 1987). Evans and Boucher (2015) argued that student autonomy has a positive impact on motivation and engagement in learning. The staff in this study identified that the SAFMEDS program allowed children to take ownership of the procedure and promoted their independence. An associated benefit of this could be scaling up the sessions (e.g., from small groups to class-wide) once children have demonstrated their ability to complete the procedure with minimal adult support.

Whilst the SAFMEDS strategy has potential in terms of scalability, teaching staff need to be wary of children cheating in larger-scale contexts. During a SAFMEDS timing, children are required to separate their cards into two piles— corrects and errors (Cihon et al., 2012). If they are recording their own data, a level of trust from the member of teaching staff is needed to ensure that they are recording the correct score independently; or perhaps intermittent supervised timings to validate each child's scores. Staff who participated in this study highlighted that some children resorted to cheating, particularly when they delivered the program on a group or class-wide scale. There are no empirical reports of children cheating during a SAFMEDS intervention within the literature. This could be due to the focus on single-case design research, whereby it would be much more difficult for cheating to go unnoticed. However, Vargas (2013) attempted to explain why children might cheat during an academic intervention using anecdotal examples. She believed that children might cheat to avoid punishment (e.g., being perceived as unintelligent in front of their peers or being verbally reprimanded by a teacher). Future research should investigate ways to reduce punishment contingencies within a SAFMEDS intervention to prevent unhealthy peer competition and help children reconceptualize low scores in a more positive way. Staff may also want to consider the advantages of using SAFMEDS on a one-on-one or small group basis in order to minimize cheating.

One method of fostering positive competition is to shift children's attention towards meeting/exceeding their own targets. Staff in this study reported that self-competition motivated children and kept them engaged with the task. Wyse (2001) outlined some key features of effective target setting. For example, it is important to set targets that are not too difficult to achieve. Staff should discuss the importance of setting targets orally to ensure children understand why they should engage with them, and children should reflect on their progress frequently. Following each SAFMEDS session, children should plot their best score on their SCC (Eshleman, 1985). The data on a child's SCC provides a visual representation of their progress over the intervention period (White, 1986). Using SCCs can be a simple and effective way of setting achievable session-by-session targets that pertain to the level of each child. Moreover, staff should consider displaying a child's overall aim for the skill on the chart (e.g., achieve 60 correct cards in 1-minute). This will help the children reflect on where they have started from and what they are aiming to achieve (White & Neely, 2012).

This study aimed to gain insight into teacher's and TA's experiences of implementing the SAMEDS strategy under the real-world conditions of the classroom. Specifically, we were interested in the perceived benefits of the strategy and any challenges that they may have faced. Gaining insight into the factors that may have affected children's outcomes, children's engagement with the strategy, fidelity of implementation, and/or persistence of implementation may help us to shape further research and training. We identified five themes relating to the social validity of using the SAFMEDS strategy to support fluency development in schools (these were: factors that promote progress, factors that limit progress, impact of competition, confidence, and inherent advantages of the SAFMEDS strategy).

Study 2: Children's views and experiences

Method

Ethics

This study received ethical approval from Bangor University's ethics committee (application number: 2018-16309). We monitored the children's assent throughout the interview. If any child indicated that they did not want to participate, the first author would have terminated the interview and informed a relevant member of staff to help the child return to their class. Likewise, if there were any questions that a child did not want to answer, the first author would not persist, she would simply move onto the next question. None of the children or their schools are identifiable within this paper to protect their anonymity.

Recruitment

We sent out an initial email to schools who had previously sent teachers/TAs to attend a SAFMEDS training session in North Wales between 2016 and 2018. Following an expression of school-level interest, we disseminated opt-in consent forms to all of the parents/guardians of children who were participating in a SAFMEDS program at the time of recruitment. This form detailed that the first author would run a one-on-one interview with their child to ask them about their experiences of the SAFMEDS strategy.

Sample

All of the children that we recruited had participated in a SAFMEDS program during the 2017-2018 academic year (October to July). This intervention focused on promoting fast and accurate recall of arithmetic facts. We received consent to interview 26 children, all of whom attended a primary school in North Wales (see Table 4.5).

Table 4.5

Demographic Characteristics of the Children who the First Author Interviewed

		n
Sex	Male	14
	Female	12
School	1	6
	2	8
	3	5
	4	4
	5	2
	6	1
School year group (age)	**2** (6-7 years)	11
	3 (7-8 years)	9
	4 (8-9 years)	5
	5 (9-10 years)	1

Procedure

The first author conducted the interviews with the children on a one-on-one basis. We opted to run a semi-structured interview for this study as opposed to focus groups for two core reasons. First, the children completed an icebreaker activity to allow them to demonstrate how they used the cards. This enabled the first author to assess if they had been correctly following the procedural steps of the SAFMEDS strategy with their teacher. It would have been difficult to collect fidelity data for every child in a group setting. Second, engaging in a one-on-one interview with the children enabled the first author to gauge their personal experience of using the strategy—reducing any social influence from their peers.

The first author conducted all of the interviews in a quiet space within the children's own school. The space was openly accessible to members of teaching staff and children within the school, but distractions were minimal. As an icebreaker activity, she asked each child if they could demonstrate how they used the SAFMEDS cards with their teacher. She provided each child the deck of skill-appropriate SAFMEDS cards (based on the skills that they had been practicing), a placemat (so they could easily separate their cards into two piles—corrects and errors), a table to record their scores, and a SCC to plot their best score onto. The children demonstrated a 1-minute SAFMEDS timing, how they engaged in an error correction procedure, and how they recorded their own data. The first author also asked the children how many timings they typically completed per session and how many times a week they came out of class for the SAFMEDS program. As the children completed this activity the first author made a note of whether they completed each of the procedural stages listed in Table 4.6.

Table 4.6

The Fidelity Checklist We Used to Establish the Extent to Which the Children Engaged with the Procedural Stages of the SAFMEDS Strategy.

		Yes/No
Before the timing		
(1)	Child shuffled the cards	
During the timing		
(2)	Child read the answer to each card out loud	
(3)	Child turned each card over to check their answers	
(4)	Child placed the cards in the corresponding pile (correct vs error)	
(5)	Child stopped after 1-minute	
After the timing		
(6)	Child counted the number of cards in their correct pile	
(7)	Child counted the number of cards in their error pile	
(8)	Child recorded their score(s) in the table	
(9)	Child recorded their best score for the session on the SCC	
(10)	Child engaged in error correction	
(11)	Child put all of their cards back into one pile	
Frequency of practice		
(12)	Child reports completing at least three timings per session	
(13)	Child reports engaging in SAFMEDS sessions at least three times per week	

Following the icebreaker activity, the first author asked the children three predetermined questions relating to their experience of using the SAFMEDS strategy. These questions focused on whether they enjoyed using the strategy, if they thought the strategy was useful, and if they use (or would like to use) the strategy outside of school. When appropriate, the first author asked the children elaborate on their answers to extract more detail or explore further themes underlying their responses. Refer to Appendix D for a full list questions and prompts that the first author asked during the semi-structured interview.

Analysis

We applied the same analysis protocol detailed in study 1 to evaluate the interview transcripts. Figure 4.2 displays the resulting thematic map.

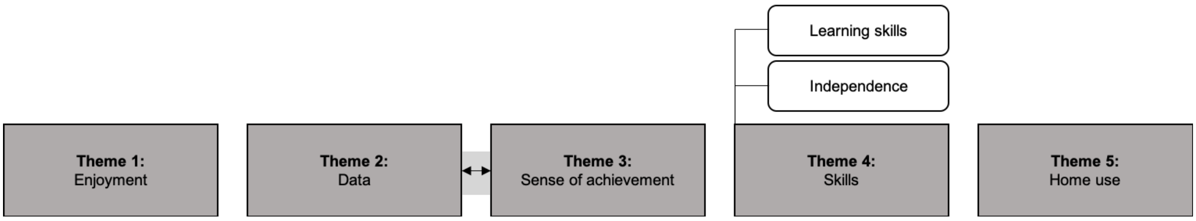

Figure 4.2. Thematic map resulting from the children's interview transcripts

Results

Fidelity

The children adhered to most of the procedural stages outlined in Table 4.6 during the icebreaker activity (M = 92.86% of stages; range: 71.43% to 100%). Table 4.7 outlines the procedural errors made by the children in each school.

Table 4.7

Percentage of SAFMEDS Steps the Children in Each School Adhered To

School	Mean % of correct procedural steps	Error (*n* of children who made each error)
1	98.81	• Did not shuffled cards (1)
2	96.43	• Did not shuffled cards (4)
3	75.71	• Did not turn each card over to check answers (2) • Child did not record their best score from the session on their SCC (5). *Note.* The teacher completed this step after their sessions.
4	96.43	• Child did not record their best score from the session on their SCC (2). *Note.* The teacher completed this step after their sessions.
5	92.86	• Child did not record their best score from the session on their SCC (2)
6	100	• N/A

Theme 1: Enjoyment

We asked the children if they enjoyed taking part in SAFMEDS sessions with their teacher. All of the children in this sample indicated that they enjoyed using the strategy. However, few were able to provide justification surrounding particular elements that they enjoyed or why they enjoyed it.

The transcripts revealed that the children saw SAFMEDS as a fun strategy to use. Two of the children also identified that they enjoyed mathematics before they started using the SAFMEDS strategy to practice their skills; which made the activity more appealing.

> [Participant 22:] ...now there is just me and [names] and that and it’s just so much fun!
>
> [Participant 1:] It's a lot of fun.
>
> [Participant 5]: (I enjoy) when you practice the cards.

Theme 2: Data

All of the children who participated in this study wrote their score down in the given table, with many of them transferring their best score for the day onto their SCC. Across the transcripts, there was a clear sense that children were motivated by counting the number of cards they answered correctly, particularly if they had exceeded their personal best score. By charting their highest score each day, the children were able to visualize their scores increasing.

> [Participant 21:] You get to go up on the chart. [KO:] you like using the chart do you? [Participant 21:] So much!

> [Participant 26:] It's enjoyable that you get better scores

> [KO:] What is your favorite part about using SAFMEDS? [Participant 13:] when we get the higher score

Theme 3: Sense of achievement

This theme is closely related to the children's attitudes towards collecting their scores but focusses on their associated perceptions and motivations. Five comments within the transcripts showed that children associated the strategy with making them more intelligent. These children used words such as "clever" and "smarter" to describe how the SAFMEDS strategy contributed to making them feel.

> [Participant 10:] (I use SAFMEDS) because I get even better at learning. I get smarter.

> [Participant 24:] Because I wasn't very good but now I'm getting better and better. My mum even sees a difference and my dad is really good at maths and he gives me some maths questions sometimes. I work them out and then most of them are all right.

The children who attended school 3 also expressed that that they were motivated by the achievement they felt as a result of their teachers' approval. They were keen to develop their fluency so that they could show her their progression. This motivated them to want to engage with further practice outside of the timetabled sessions within school.

> [KO:] Would you like to use it outside of school if you could? [Participant 23:] *nods* [KO:] Why would you like to do that? [Participant 23:] So I could show Miss [name] how well I did.

One way in which the children felt that they were showing improvement was by answering more cards correctly than they had achieved on previous attempts. As such, several comments alluded to the achievement that the children felt when they were able to answer all of the cards in the pack correctly within 1-minute. Also, these children recognized that since taking part in the SAFMEDS program they had become more fluent at recalling arithmetic facts. Achieving fluency appeared to encourage children to engage with the strategy, both in and outside of school.

> [KO:] why would you like to do (the SAFMEDS strategy) at home? [Participant 1:] Because I can learn better when I go to school. I can do it will get more faster.

Theme 4: Skills

All of the children alluded to the skills that they had developed throughout the course of their involvement with the SAFMEDS program. We identified two sub-categories relating to skill development: learning skills and independence.

Learning skills

We asked the children if and why they thought the SAFMEDS strategy was useful. Many of the children identified that the strategy helped them to improve their mathematics skills, including supporting their fluency of arithmetic facts. Some children went on to explain that they had worked through multiple packs across their involvement with their SAFMEDS program, as such they were able to answer progressively more difficult mathematics questions.

> [KO:] yeah? What do you enjoy about (using the SAFMEDS strategy)? [Participant 11:] It helps me with maths.

> [KO:] okay, what is your favorite part about doing the cards? [Participant 2:] Where I swap over cards. [KO:] Do you mean when you move onto different packs? [Participant 2:] yeah.

> [KO:] What do you enjoy about (using the SAFMEDS strategy)? [Participant 24:] Well, last year I wasn't good at maths because I wasn't very good at like adding stuff and taking away. But I … (teachers name) came to collect me … she told me what I was doing and then we went on to the adds … then, like a couple of months later I went onto the take aways and then I went onto the adds and take aways.

Independence

Aside from the academic gains that come from using the SAFMEDS strategy, a prevalent theme amongst the transcripts was the independence that the strategy offers children. The majority of children we interviewed were keen to use the cards independently, even if support from an adult was accessible (such as from a parent).

> [KO:] Would you use them on your own or would you use them with your mum and dad? [Participant 7:] On my own.

> [KO:] Do you do them on your own or with your parents? [Participant 10:] I do it on my own because I have a timer on my PlayStation.

> [KO:] What's your favorite part about doing SAFMEDS? [Participant 17:] That you get to fill in your own sheet.

Theme 5: Home use

Eight children claimed that they already use the cards to practice using the SAFMEDS strategy at home. A further 14 children identified that they would take the cards home to practice if they were given the option to. The children who participated in these interviews saw the advantages of practicing the cards regularly; as it helped them to develop their mathematics skills at a faster pace than just practicing at school.

> [Participant 2:] Is there an app on the computer you can get (SAFMEDS) on? [KO:] No, not yet. Would you like an app? [Participant 2:] Yeah. [KO:] Why would you like an app? [Participant 2:] Because I can use it outside of school and in school.

> [KO:] Do you use SAFMEDS at home? [Participant 15:] *nods*. [KO:] Why do you use them at home? [Participant 15:] To get better at this.

> [KO:] Why would you like to use (SAFMEDS) at home? [Participant 13:] So then I learn every day.

Whilst the majority of the children we interviewed claimed to want to use the cards independently, eight children identified that they would like to involve their parents in the strategy if they could take the cards home. Examples of support parents could offer included holding their child's cards during the timing and helping their child to address any cards in their "not yet" pile (via error correction).

> [KO:] Do you like having your parents helping you (to do the SAFMEDS strategy)? [Participant 5:] Yeah. [KO:] Why? [Participant 5:] Because it helps. [KO:] How do they help you? What do they do? [Participant 5:] They hold the cards and then I said the answer and they put it down.

> [KO:] Would you do (SAFMEDS) by yourself or with someone else? [Participant 24:] I would do them with somebody else, but they would give me the timer and then I would make my own board, probably. Then, my mum will probably time me on the timer. When it's stopped, I will probably tell my mum and then she'll do my not yets.

Three children suggested that other activities available to them around the house reduce the appeal of using the SAFMEDS strategy at home. These children still claimed to enjoy using the strategy with their teacher but would opt to leave the flashcards at school.

> [KO:] Would you like to use them at home more? [Participant 4:] No. [KO:] No? Why not? [Participant 4:] Because we have loads of stuff to do at our house.

Discussion

Previous research has demonstrated that the SAFMEDS strategy can help children who attend mainstream schools to improve their mathematics skills (e.g., Greene et al., 2018; Hunter et al., 2016). However, no known research has robustly evaluated the social validity of the strategy from the perspective of the children who have used it. This study aimed to provide some insight into primary-aged children's experiences of using the SAFMEDS strategy with their teachers in their school. We were successful in identifying five themes; each demonstrating some benefits of using the SAFMEDS strategy to support development of mathematics skills. Specifically, these themes provided a broad insight into factors that might affect children's engagement with the strategy.

Hunter et al. (2016) highlighted the need for research assessing the social validity of the SAFMEDS strategy to provide evidence of its acceptability to aid the development of mathematics skills. All of the children we interviewed for the current study indicated that they enjoyed using the strategy, with many also indicating that they would engage in further practice at home if they were able to. Moreover, the children were able to identify the progress that they have made across their engagement with the SAFMEDS program; with particular emphasis on their developing mathematics skills and independence. This result provides support for the argument that children associate a level of social validity with using the SAFMEDS strategy.

The children we interviewed for this study demonstrated a preference for independent practice. This has advantages with regards to scaling up the intervention and promoting a level of autonomy within the classroom. Previous research provides evidence to suggest that practitioners can use the SAFMEDS strategy to support mathematics development on a single-case (e.g., Casey et al., 2003; Cunningham et al., 2012), small group (e.g., Beverley, Hughes, & Hastings, 2018), or class-wide basis (Hunter et al., 2016). Taken together with the results from study 1, our findings suggest children are able to take ownership of the SAFMEDS strategy, which in turn promotes independence and establishes a routine over the course of the school year. An additional advantage of an intervention promoting high levels of independence is that children can engage with it in any environment that they choose to (e.g., in school or at home). There is currently a lack of published research reporting the effects of children using the SAFMEDS strategy at home to support their learning. Investigating this further would help us to validate children's willingness to engage with the

SAFMEDS strategy outside of school and evaluate the educational effects of the additional practice.

Some children did identify that they would like some support from an adult whilst they completed the strategy. Examples included providing support during error correction and someone manipulating the cards for them during the SAFMEDS timing. Whilst providing additional instruction to teach new skills during error correction is an imperative part of the learning process, practitioners should be wary of providing too much support during the SAFMEDS timing itself—including setting the pace at which the children can answer the cards. Binder (2003) explained that certain teaching methods can prevent children from progressing skills at their own pace. As practitioners within education (e.g., researchers and teachers), we should design programs that lift ceilings on children's performance to enable them to achieve levels of fluency that ensure retention and application to other skills (Binder, 1996). Lindsley (1996b) explained that a learner's rate of responding decreases by approximately half if someone else holds and directs the cards for them, compared to self-held performance. This is due to learners being able to present the next card to themselves twice as fast as a partner could. The results from our study suggest that most children enjoy the independence that the SAFMEDS strategy can offer them, so it is worthwhile exploring how to harness this within future training and research.

A cardinal feature of the PT approach is the using the SCC to plot and visualize data (Calkin, 2005). Using the SCC allows for assessment and intervention to occur concurrently; making it a useful tool for practitioners to make real-time, data-based decisions (Aninao, Acevedo, Newsome, & Newsome, 2015). From a child's perspective, the chart provides a way of visualizing their progress. Johnson and Street (2012) explained that the trajectory of data on the SCC can help practitioners and/or children to predict what their score will be for the subsequent sessions. In the context of improving learning pictures, this can be a simple and effective practice to help set attainable score targets on a child-by-child basis for each session. The children we interviewed emphasized the motivational power that data can have on their engagement with the SAFMEDS strategy; particularly with regards to setting themselves a target each session (i.e., beating their personal best score for the pack of cards they are working on). Applying a pragmatic approach to goal setting using trajectory lines might help keep children focused and working towards attainable targets.

Within this study we interviewed primary school children. However, children's views on the strategy might vary depending on their age and the educational institute that they attend (e.g., primary school, secondary school, higher education, or further education). A valid extension of this research might investigate learners' views on the SAFMEDS strategy at different stages of the education continuum. For example, we could consider why secondary school children might choose to use SAFMEDS as a revision strategy for their pending statutory exams (e.g., GSCEs, A-Levels). This extension would provide a more holistic overview of children's views of the SAFMEDS strategy.

We identified five themes relating to children's experiences of using the SAFMEDS strategy to support their skill development. These themes were: enjoyment, data, sense of achievement, skills, and home use. At a broad level, these themes each provide insight into factors that may affect primary school children's engagement with the SAFMEDS strategy.

General discussion and conclusions

By gaining insight into how the SAFMEDS strategy works under the day-to-day conditions of the classroom, we can better understand how to support practitioners (including teaching staff and researchers) with any unanticipated challenges that they may face. This may contribute towards supporting practitioners to persist with the implementation of the strategy and maintain/improve children's engagement. Moreover, we can reflect upon the positive elements associated with using the strategy to ensure that we harness them in training and practice.

The children who participated in this research reported that they enjoy using the SAFMEDS strategy improve their basic mathematics skills. This opinion was validated by teaching staff who noted the visible enthusiasm that their children displayed when coming out of class to engage with the SAFMEDS sessions. This reflects the social validity of the strategy from the perspective of the children who engage with the strategy regularly, with many also indicating that they would use it outside of school if they could.

Aside from the academic skill development, the children we interviewed showed an appreciation for the independence that the strategy afforded them. This is advantageous in terms of scaling up the program and considering when and where children could engage with the strategy. Several members of staff were keen to scale-up the intervention beyond small groups to help support more children in their school. They felt that this would be attainable once the children were proficient enough to engage with the strategy independently. This demonstrates the suitability of the SAFMEDS strategy as a tool for assessing skills and developing fluency in educational contexts. However, teaching staff should be wary of cheating tactics that the children might engage in when scaling up the program.

One way to help reduce cheating within larger groups might be to set the children personalized and attainable goals for each session. Staff reported that this tactic shifts the focus away from unhealthy peer-competition towards self-competition. The children we interviewed reiterated that they were motivated by beating their personal best scores each session. By making their progress visual on a SCC, it might be possible to set attainable goals based on their performance across previous sessions. Moreover, staff and children can use this data to reflect on their progress for specific mathematics skills.

The SAFMEDS strategy is designed to help learners develop fluency and highlight areas where they might need support (e.g., additional instruction or adapted materials). The teaching staff and the children who participated in this research were able to acknowledge that the SAFMEDS strategy aids fast and accurate recall of arithmetic facts. Many of the children also progressed onto more difficult mathematics skills throughout their involvement in a SAFMEDS program.

Chapter 5 : Supporting literacy and numeracy wit hin Inner-City London: A charity led fluency-building program

Project collaborators: Owen, K., Illot, N., Hunter, S., Aylward, R., Payne, J., Marchant, S., & Hughes, J. C.

Preface

As a charity, XLP (n.d.b) aim to support children from areas of high social deprivation to complete their education and avoid engaging in anti-social acts within the community. A member of the XLP team had heard about the SAFMEDS strategy and what it can achieve, so was interested in how this could be adopted within educational provisions across inner city London to support the children that they work with. This chapter provides an example of a close-to-practice impact report summarizing the process of collaboration between researchers, a charity, and educational provisions as well as the literacy and numeracy progress of the children that we worked with. This was an exploratory piece of research to investigate the impact of the pilot of the XL-LAN (literacy and numeracy) project. As such, the data reflects the children's typical in-session SAFMEDS progress and the social validity that they associated with the strategy. XLP passed on the findings of this research onto their funders to show (1) the impact that this program could offer and (2) recommendations for improvements to the scheme going forward. We discuss some further implications, applications, and limitations of this research in Chapter 7.

Abstract

The XLP numeracy and literacy (XL-LAN) project aims to support children who are unable to perform age-expected literacy and/or numeracy skills fluently. XLP recruited the expertise of researchers at Bangor University to help train their staff in the use of the Say-All-Fast-Minute-Every-Day-Shuffled (SAFMEDS) strategy. XLP and researchers at Bangor University continued to collaborate in order to support the implementation of the program in educational environments across inner city London; this includes mainstream schools, pupil referral units, and mentoring services. In this report, we outline the process of this collaboration and the findings from the XL-LAN pilot project. Using a mixed-method approach, we were able to identify key quantitative differences relating to the children's in-session SAFMEDS scores, as well as qualitative themes relating to the children's experience of using the strategy to support their learning. The final dataset included quantitative data from 263 children. Children's mean fluency scores appeared to improve irrespective of the subject they were practicing, however progress appeared to be more stable and consistent when practicing literacy skills. Interviews with 38 children who had used the SAFMEDS strategy revealed five themes: (1) procedure, (2) improvement, (3) revision tool, (4) home use, and (5) withdrawal from class. Collectively, these data provide useful feedback which may shape the future delivery of the XL-LAN project and other projects utilizing the SAFMEDS strategy in similar contexts.

The Programme for International Student Assessment (PISA) identifies the extent to which 15-year-old students have gained skills that are essential to function independently in society (OECD, 2016a). Jerrim and Wyness (2016) benchmarked London in the PISA reports by predicting London's mathematics and reading results across the 2009 and 2015 assessments. London's reading scores placed them 26th out of the 37 economies included in the analysis. This result is significantly below other major European cities. The mathematics averages revealed that students in London fall significantly below 22 other economies.

Banerjee's (2016) systematic review revealed several global factors that can predict lower levels of childhood attainment. These include socio-economic status, ethnic minority status, speaking English as an additional language, and immigrant status. London is made up of a demographically diverse community. In inner city London approximately half of school children do not speak English as a first language (Department for Education, 2016c). Demie and Strand (2006) found that fluency in the English Language is positively correlated with the number of A*-C grade GCSEs achieved in London schools. Underachievement at school impacts individuals and the wider society; making it an important problem to target within education. Wolf (2011) associated underachievement at school with limited job prospects and fewer opportunities with regards to further education. Moreover, Henry, Knight, and Thornberry (2012) explained that the frustration children can feel as a result of failing at school can lead to truanting, drug use, and engagement in criminal activity.

These statistics provide the context for the close-to-practice project we present within this report. There is clear need to identify effective educational strategies that can raise English (literacy) and mathematics (numeracy) outcomes of children across London. By identifying remedial interventions, we may be able to target groups who are proportionally more at risk of poor academic outcomes and limit some of the negative consequences associated with underachievement.

XLP (n.d.b) are a youth charity who provide support and services for school-aged children across nine inner-city boroughs around London. Their mission is to empower children who are at risk of academic underachievement to complete their education, refrain from engaging in anti-social behavior, and positively contribute to their community. In 2016, XLP (n.d.c) initiated a pilot of their literacy and numeracy (XL-LAN) project, which focused on promoting basic skills across the English (literacy) and mathematics curriculum in London. This scheme aimed to help children improve the speed and accuracy (fluency) in which they can perform basic skills using the Say-All-Fast-Minute-Every-Day-Shuffled (SAFMEDS) strategy.

SAFMEDS is a practice and assessment technique derived from the principles of precision teaching (PT; Potts, Eshleman, & Cooper, 1993). The guiding principle of PT is that we should monitor skill fluency regularly and chart progress. Charted data gives us the ability to assess when children need additional support in real time, thus helping children to access support when necessary and preventing prolonged periods of non-progression (Merbitz, Vieitez, Merbitz, & Pennypacker, 2004). Children typically engage with the SAFMEDS strategy through a deck of flashcards. On the front of each card is a question or statement and on the back is the corresponding correct answer. A child aims to get through as many cards as they can within 1-minute (Cihon, Stutrz, & Eshleman, 2012). After each session the child should plot their best score on a standard celeration chart (SCC; see White & Neely, 2012). Over time, the SCC depicts learning pictures (Lindsley, 1995). These pictures allow children and teachers to decided when additional support is necessary. Ultimately, the SAFMEDS strategy promotes fluency and identifies any gaps in a child’s knowledge.

Over several years researchers at Bangor University have been investigating the effects of the SAFMEDS strategy in education. Collectively this research has demonstrated the utility of the SAFMEDS strategy in mainstream primary schools to build fluency in basic mathematics skills (Hunter, Beverley, Parkinson, & Hughes, 2016) and second-language vocabulary (Beverley, Hughes, & Hastings, 2016). Moreover, Hunter, Geary, and Hughes (2018) demonstrated that older learners who are eligible to sit statutory exams (i.e., A Levels) can use the strategy to consolidate knowledge more effectively than mind-mapping. Despite a growing evidence-base for SAFMEDS research, few studies have evaluated the effects of the strategy when experimental control is limited. Owen et al. (under review) found that a researcher can coach teachers to implement the SAFMEDS strategy in their school. The combination of initial training and individualized implementation support visits helped teachers to elicit greater fluency gains from the children that they worked with, comparative to teachers who only received the initial training. This provided the foundation to explore additional ways of providing support to schools to implement the strategy.

During the pilot of the XL-LAN project, XLP enlisted the help of researchers at Bangor University. The collaboration between these two organizations aimed to satisfy two desirable outcomes. First, the project could benefit from the ongoing trials happening in North Wales. Second, the data produced from the XL-LAN project could expand on our knowledge of how the SAFMEDS strategy can be used in environments where experimental control is not present. In this paper, we present the quantitative (study 1) and qualitative (study 2) findings from the 2-year pilot of the XL-LAN project. We were interested in assessing provisions' engagement in the program, the fluency progress of the children, and the social validity children associate with using the SAFMEDS strategy.

Study 1: A quantitative exploration of the XL-LAN data

Method

Collaboration

One advantage of close-to-practice research is that key stakeholders can share ideas and expertise. In this example, researchers at Bangor University were able to offer advice and expertise in relation to the theory underlying the SAFMEDS strategy and analysis of the XL-LAN data. They were also able to share findings from trials using the SAFMEDS strategy in schools across North Wales, as they were happening. The staff at XLP had clear objectives about who they wanted to support (e.g., children enrolled on their mentoring schemes, those who use their youth bus service, and children attending educational establishments in inner city London). As the project progressed, it was the staff at XLP who were out in the field maintaining and supporting the XL-LAN SAFMEDS program. They were able to report the progress and challenges as they occurred to the first author. Below we discuss some of the key features of the collaboration model.

Training

At the foundation of this model lies training. After some initial discussion we opted to employ a ‘train the trainer’ approach. That is, a researcher with experience using the SAFMEDS strategy in schools trained the staff at XLP to deliver SAFMEDS sessions. This enabled the XLP staff to go forward and replicate the training for each of the educational stakeholders they were working with. The training for this project mirrored the training detailed in Chapters 3 of this thesis, with the added emphasis on some of the key features of the program. We advised that: (1) supervisors should encourage children to get through as many cards as they can in 1-minute to increase fluency and highlight error correction opportunities (i.e., getting faster); (2) where appropriate, children should engage with the strategy independently (e.g., hold and direct their own cards, chart their own data) and; (3) supervisors should set attainable and individualized score targets for each child each session.

Creating materials

Researchers at Bangor University had a pre-existing bank of flashcard stimuli to facilitate the SAFMEDS strategy. The packs covered some of the core skills prescribed by the Key Stage 1 and Key Stage 2 numeracy curriculum in Wales. Staff at XLP expanded upon these resources to create packs that supported the literacy and numeracy curriculum in England. XLP created a catalogue of 57 literacy and 63 numeracy packs, with each pack containing 60 cards. These packs supported the skills that children are expected to have mastered by the end of primary school (aged 11-years). By building fluency in these basic skills, XLP hoped to address some of the skill deficits experienced by the children that they support. A researcher at Bangor University validated each pack to ensure that it assessed each skill appropriately (e.g., the cards did not contain too many words or irrelevant cues).

Consultancy

In-house and phone progress meetings provided a platform for us to update each other on the project. If the XLP staff had encountered any challenges during school visits, this contact offered an opportunity to receive advice and suggestions for forthcoming practice from the first author who had experience in the field. At the end of each month, a member of staff at XLP sent project data to the first author for review. This provided an opportunity to identify any in-session SAFMEDS scores that required attention. For example, there might have been reason to believe a child had been cheating (e.g., inflated scores) or not progressing desirably (e.g., plateauing or worsening scores). We discussed these instances and possible ways to overcome the associated challenges. Where appropriate, a member of staff at XLP relayed this feedback to the corresponding supervisors in the education provisions.

Analysis and evaluation

In order to assess the impact of the XL-LAN project, researchers at Bangor University analyzed the data that XLP had collated throughout the pilot period. It was important for us to identify strengths and areas for improvement within this project in order to enable XLP to adapt and enhance the service that they offer. Researchers at Bangor University analyzed this data, summarized the outcomes, and offered recommendations for future practice (as outlined in the discussion).

The XL-LAN pilot

During the first year of the project, XLP trained the supervisors in each education provision to use the strategy with the children that they work with. This included training for XLP staff and mentors, teaching staff in mainstream schools and pupil referral units (PRUs), as well as staff working for external youth projects. Following training, XLP provided each supervisor with an online training manual, which they could refer to throughout their involvement with the XL-LAN project. This manual included a video modelling the strategy, instructions on how to chart data with reference to learning pictures, and tips for successful implementation (e.g., how to address cheating). Moreover, this manual contained a list of all the SAFMEDS packs available and the corresponding level of mastery for each pack (i.e., the minimum number of correct cards per minute a child would be expected to achieve before moving onto the next progressive skill). Staff at XLP continued to support provisions with their implementation of the strategy following training through in-house visits, phone calls, and email contact.

Stakeholder roles

Researchers at Bangor University supported the project at a consultancy-level. This including the initial training for the XL-LAN team, answering any project-specific questions raised by the XL-LAN coordinators, analysis of the project data, and providing recommendations for future practice. XLP replicated the training they received for the supervisors delivering the SAFMEDS program, they also distributed the resources and facilitated regular contact with each of the supervisors to support implementation and progression. The supervisors in each provision were expected to accommodate the program and support the children that they work with through the SAFMEDS packs. This included reviewing the children's data regularly and providing additional support to teach skills when necessary.

Ethical considerations

At the beginning of the project, XLP asked all participating provisions to sign a partnership agreement. This outlined what provisions and supervisors could expect from XLP throughout the duration the project. XLP agreed that they would provide training, resources, and support throughout; whilst provisions were expected to provide XLP with the student data needed for analysis. In addition to this, we provided each provision with an information sheet about the project. Each provision had to provide written consent to enable researchers at Bangor University to analyze their anonymized data.

Sample

XLP operate in nine boroughs across inner city London; supporting children in several settings. This pilot contains data from seven secondary schools, three primary schools, one pupil referral unit, a mentoring network, and a tutoring agency. The database included data from 551 active children, each registered to at least one pack of SAFMEDS cards. However, not all of these data were included in the final analysis.

Data inclusion criteria

When collecting data, supervisors were able to note down if they observed the children for a complete SAFMEDS timing (from shuffling their cards to writing down their score) and could validate their score. Observed timings provided a reference point for progression. If a children's scores deviated too far from the observed session, then supervisors or XL-LAN coordinators could question the validity of these values. We excluded any data points that were deemed invalid by supervisors or staff at XLP. We also excluded the data from children who had completed an insufficient number of SAFMEDS sessions. The final dataset included data from 263 children (122 females, 141 males). Each child had completed at least seven SAFMEDS sessions, over a period of 3 consecutive months. Table 5.1 outlines some further demographic characteristics of our sample.

Table 5.1

The Number of Children Registered to Valid SAFMEDS Data, Categorized by Their Corresponding School Year Group

Year group	Age range (years)	*n*
3	7-8	23
4	8-9	26
5	9-10	54
6	10-11	28
7	11-12	73
8	12-13	30
9	13-14	24
10	14-15	3
11	15-16	2

Results

Pack usage

Throughout the duration of the project, XLP issued 726 packs of cards. Several of the children were registered to multiple packs, due to skill progression. We included data from 363 of these packs in our final analysis (as per the inclusion criteria). A total of 201 children completed numeracy packs and 64 children completed literacy packs. Descriptive statistics revealed that few children used more than one SAFMEDS pack throughout the duration of the program. These details are listed in Table 5.2.

Table 5.2

Number of Literacy and Numeracy Packs Completed by the Children

Total number of packs used	Number of children	
	Literacy	Numeracy
1	57	119
2	6	69
3	-	10
5	1	1
7	-	1
9	-	1

Number of sessions

Table 5.3 displays the median number of sessions completed by children on each SAFMEDS pack. It should be noted that this data does not necessary reflect the number of sessions needed to achieve mastery. Instead, this data provides some insight into provisions' engagement with the XL-LAN program.

Table 5.3

Median Number of Sessions Spent on Each SAFMEDS Pack and the Frequency of Occurrence of These Sessions (Per Week)

	Total number of sessions	Number of sessions per week
Literacy	6	2
	[*IQR* = 7, min = 1, max = 35]	[*IQR* = 2, min = 1, max = 5]
Numeracy	6.5	2
	[*IQR* = 6.75, min = 1, max = 26]	[*IQR* = 1, min = 1, max = 6]

Note. IQR = Inter-quartile range.

Performance statistics

We calculated a mean slope for children's performance on their first pack of SAFMEDS cards. This calculation took into consideration each child's score during their first SAFMEDS session (S1), their personal best score (PB), and the number of sessions it took to reach their PB (NS) [Calculation: (PB – S1)/NS]. The data shows that children improved by a mean of 1.97 correct cards per session for the numeracy packs (*SD* = 1.22); and 2.48 correct cards per session for the literacy packs (*SD* = 1.56).

Due to the disparity in the children's baseline scores during S1, and therefore the progress that they could make before they met the mastery criteria, we have also reported the number of cards that they improved by (PB – S1). This data is displayed in Table 5.4. It should be noted that the number of sessions completed by children varied. Moreover, there was some variation within the time between sessions due factors such as school half terms and illness.

Table 5.4

Score improvement Between Children's Session 1 (S1) Correct Score and their Personal Best (PB)

		Score improvement between S1 and PB				
		0-10	11-20	21-30	31-40	41-50
Number of children	Literacy	15	22	12	12	3
	Numeracy	59	87	45	9	1

Figure 5.1 depicts the children's mean correct scores on the numeracy and literacy packs across sessions. The numeracy data shows that the number of correct cards the children answered increased steadily between sessions 1-9. After session 9 this progress is visibly more variable. Few children were successful in achieving a correct score that would typically reflect that they have mastered the practiced skill (i.e., 40-60 correct cards in 1-min). It is important to note that, for the most part, numeracy performance is relatively stable. The

decrease in performance in sessions 22 to 26 is likely due to the small sample size included in the calculation causing more natural variance in scores.

The literacy data in Figure 5.1 shows that the number of correct cards children answered increases steadily across all sessions. However, again this progress is visibly more variable as the sample size decreases. It is not possible to make inferences from this data with regards to the number of children who achieved mastery on literacy packs because the mastery criteria varied between 20-30 corrects and 40-60 correct per minute, depending on the skill the children practiced. Across both literacy and numeracy packs it is clear that as session number increased so did the number of children who completed the sessions. This is in part due to provisions' engagement with the program and also when the provision enrolled onto the project (some started before others).

Figure 5.2 reflects the progress made by children between their first session and their personal best score. Here we focused on their first pack of cards because this represents data from the whole sample and this is typically the pack that we would expect them to make the most progress on during the program (i.e., the children should have mastered the first pack before moving on to another). To calculate proportions, we took the children's highest score during their first session and divided it by the total number of cards in the pack (S1/60). We repeated this calculation for the children's personal best score for that pack (PB/60). During S1, the children were able to answer an average of 26% of the numeracy cards in the pack correctly in 1-minute. This increased to an average of 52% of the pack on their PB try. The literacy scores increased from 13% of the deck during S1 to 47% during their PB try. These differences were statistically significant and robust, demonstrating a strong, consistent pattern of improvement for most children [Welch's $\text{t-test}_{\text{Numeracy}}$: $t(200) = 26.59, p < .0001, d = 1.82$; Welch's $\text{t-test}_{\text{Literacy}}$: $t(63) = 13.19, p < .0001, d = 1.65$].

Provision/supervisor effects

We collapsed the children's mean scores per session to evaluate the different in progress across different provisions/supervisors (see Figure 5.3). These data may suggest that some supervisors are able to elicit steeper positive fluency gains from their children than others. This could be due to factors such employing the SAFMEDS strategy more regularly, and/or following the stages more accurately; although at this stage we have no evidence to confirm this. It is also plausible that these effects lie at a provision level, whereby the children attending certain provisions might exhibit differing behavioral traits (i.e., are more likely to be disengaged or engage in behaviors that challenge). Similarly, certain demographic groups might be more likely to progress at a slower trajectory than their peers (e.g., children from low socio-economic households, children with additional learning needs, or children who speak English as a second language). Access to demographic data such as this would help us to predict which factors influence performance growth.

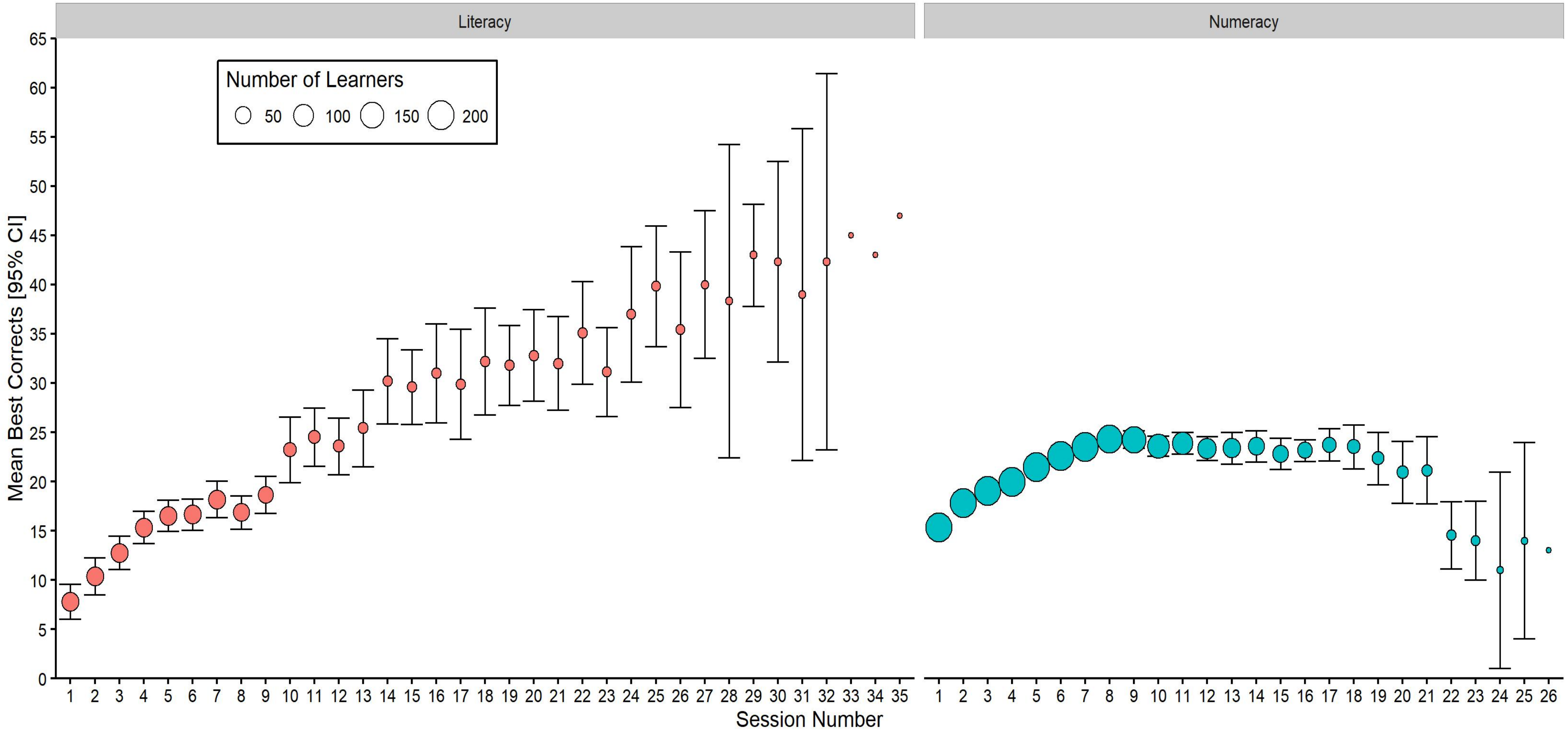

Figure 5.1. Mean score progress across SAFMEDS sessions for the literacy and numeracy packs. The colored bubbles represent the number of children who completed *n* number of sessions; with larger bubbles representing more children. Error bars represent 95% confidence intervals.

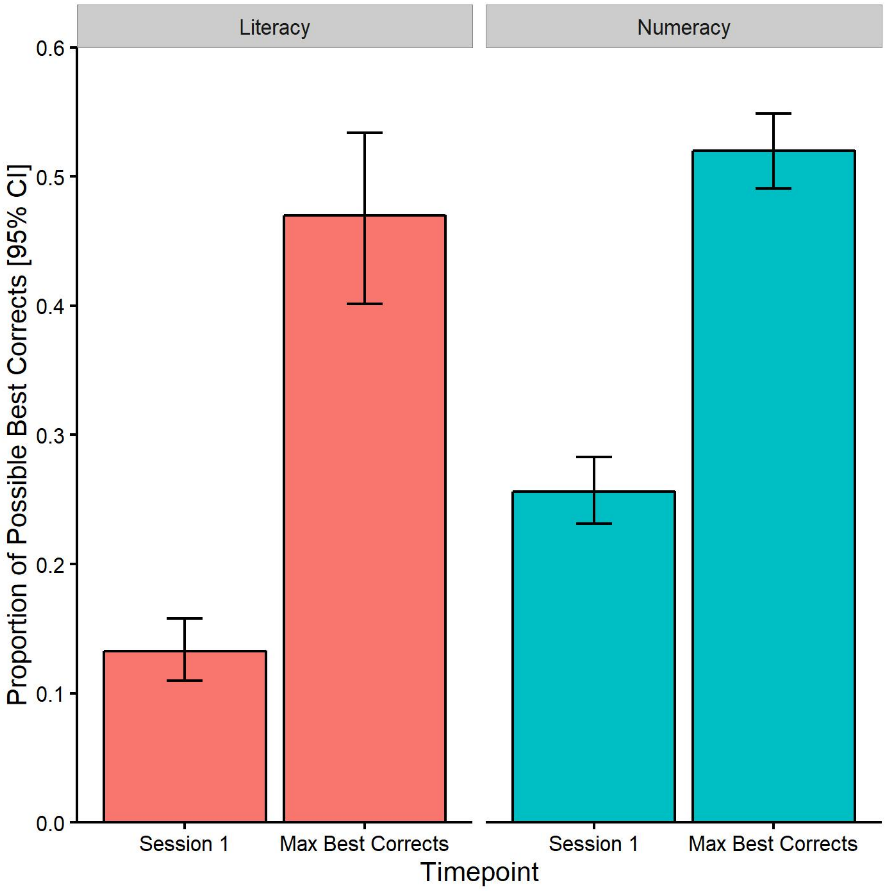

Figure 5.2. Mean proportion of best correct responses (out of 60 possible cards) for session 1, compared to their personal best across all sessions. This data reflects the first pack of cards each child attempted. Error bars represent 95% confidence intervals.

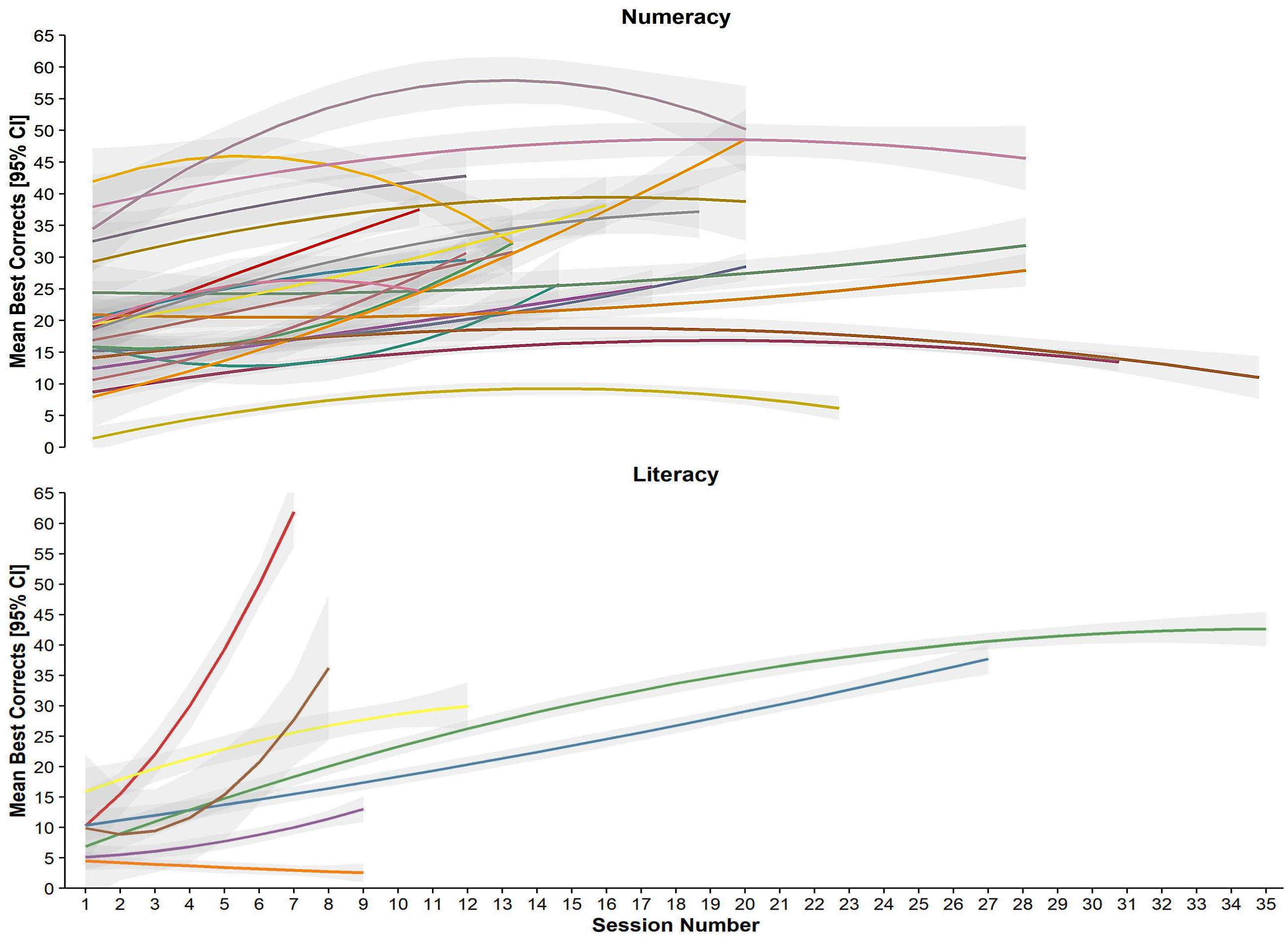

Figure 5.3. Estimates of learning slopes for each provision/supervisor. Each colored line represents the collapsed mean score for per session for their first pack of cards. Grey bands represent 95% confidence intervals. *Note*. Estimates after sessions 12 are likely to be unreliable as number of children reduced substantially.

Discussion

Since 2016, the XL-LAN project has supported the learning of children across inner city London. The data that we have presented within study 1 provides insight into how children's fluency of key literacy and numeracy skills may *typically* progress if they use the SAFMEDS strategy. The numeracy data suggests that children make steady progress within the first nine SAFMEDS sessions before their scores becomes more variable; whilst the literacy data suggests that children's scores steadily and consistently increase throughout a SAFMEDS program. Our data also suggests that the children can answer significantly more cards in each pack correctly after taking part in the XL-LAN program. Moreover, nearly all of the provision/supervisor-level data shows an increase in children scores over time. Together, this data supports the effectiveness of the SAFMEDS strategy in education environments, even when there is limited control over confounding variables. In this section, we discuss some of the factors that may have influenced the variability within the data to help conceptualize some next steps for the XL-LAN project. Whilst these suggestions relate to the XL-LAN project, they may also be useful considerations for other school-based SAFMEDS projects with similar samples (i.e., children at risk of academic failure).

The data revealed that learning increased at a faster rate for the literacy packs than the numeracy packs. Lortie-Forgues, Tian, and Siegler (2015) argued that children use different strategies to approach the same numerical question and complex numeracy skills require children to handle numbers differently. This adds complexity to the way that children answer numeracy questions and how educators may teach these skills. These types of problems lend themselves to the PT approach, because charting data regularly allows us to see when children need additional support. For example, if a child's data shows that they are making little or no progress, they may need their teacher to explain a concept or computation method in a different way before they can master the skill (Lindsley,1995). Charting data using the

SCC is one of the fundamental principles that guide PT approaches (Lindsley, 1995). Without the interpretation of data, practitioners compromise the fidelity of a SAFMEDS program, and children are likely to show limited progression (White, 2000). It is possible that some of the supervisors involved in the XL-LAN project did not use the data that they collected to make decisions about when to intervene. This might explain why the numeracy data plateaus below the level of expected threshold to demonstrate mastery (≥ 40 correct cards in 1-minute) after 12 sessions. It is worth noting that this argument may also apply to the literacy data. However, the mastery criteria for different literacy packs varied (from ≥25+ to ≥40+ correct cards in 1-minute) so it is not possible to draw this conclusion from the collapsed data. Further research is necessary to establish the extent to which supervisors use the SCC to facilitate data-driven decision. If they do not, then it may be worth exploring some of the underlying mechanisms that may affect the use of the SCC in schools (e.g., lack of time, perceived unimportance, competing methods of recording data).

The trend of the numeracy data may also be explained by the regularity of SAFMEDS practice. For some children, their engagement in SAFMEDS sessions was sporadic; with prolonged periods of time between sessions. This is an applied problem that is common in education environments due to factors such as poor attendance, school holidays, and internal school events (Mason, Rivera, & Arriaga, 2018; Tincani, 2004). However, the SAFMEDS approach relies on regular practice to ensure that children are able to contact the material, address any “not yets”, and elicit retention (Graf & Auman, 2005). Without this regularity, the strategy is less likely to have the desired effect. The data from this study showed that children completed an average of two SAFMEDS sessions per week; which is one-day fewer than employed within existing research (e.g., Greene, Mc Tiernan, & Holloway, 2018; Hunter et al., 2016). This highlights a need to consider strategies to integrate SAFMEDS more readily into the school timetable to ensure that children have the opportunity to practice and assess their skills regularly.

The confidence intervals and spread of data within our analyses show us that there is lots of variability between children's scores. The cause of this variability is unclear and is likely to be multi-faceted. Several factors such as, additional learning needs, first language, and socio-economic status may have affected children's performance and progress. Previous SAFMEDS literature favors the use of single-case research designs; whereby researchers do not need to account for the variance caused individual differences (e.g., Casey, McLaughlin, Weber, & Everson, 2003; Munizi & McLaughlin, 2013). This could explain some of the differences in group means that we have reported compared to the individual progress of children in the available published research. Additional demographic data may help us predict which variables affect SAFMEDS performance. XLP did request this information from each provision, however due to low return rates we were unable to draw robust conclusions from our dataset.

Variability within our dataset may extend to the delivery of the program across different provisions; however, we need further data relating to demographics and the fidelity of implementation (i.e., how closely supervisors and children implemented the program in line with its intended design) to assess this assumption. Going forward, it would be useful to use observational checklists to assess the extent to which supervisors and children adhere to features prescribed within the program (Schoenwald et al., 2010). Staff at XLP could also use these checklists to provide individualized feedback to supervisors with regards to the positive aspects of their implementation and any recommended changes that they should make.

Study 2: Evaluating the social validity of the XL-LAN project

Method

Sample

Staff at XLP recruited 38 children who had used the SAFMEDS strategy as part of the XL-LAN pilot to participate in an interview. The sample included children from four different schools (1 primary school; 3 secondary schools). The children were aged between 6 and 15 years old. Refer to Table 5.5 for further demographic information.

Table 5.5

Demographic Characteristics of the Children who Participated in the XL-LAN Interviews

		Sex (*n*)	
Year group	Age (years)	Male	Female
3	7-8	2	3
4	8-9	1	4
7	11-12	7	4
9	12-13	10	4
10	14-15	1	2

Interview format

Two members of staff from XLP conducted the interviews on either a one-on-one or small group basis; depending on the preference of the schools and children. They asked the children a set of predetermined questions (see Appendix D). Where appropriate, the XL-LAN staff asked for expansion or follow-up questions.

Data analysis

For this study, like in Chapter 2, the first author adhered the six stages of thematic analysis outline by Braun and Clarke (2006). One of the XL-LAN project coordinators transcribed each of the interviews before emailing the completed transcripts to the first author for analysis. The first author became familiar with the content of the interview transcripts by reading each of them through several times. After reading through them twice without comment, the first author began to make some initial notes about interesting and salient comments. The first author later manually coded the key features of each data extract (for example see Table 5.6); each of these codes reflected a feature of the data that she could assess in a meaningful way. After coding the data, the first author transferred each code onto post-it notes to help assist with the identification of potential themes. After refining and naming each theme, the first author had generated a thematic map depicting how each of the themes related to one and other (see Figure 5.4). Each distinct theme encapsulates salient groups of comments within the transcripts. One of the XL-LAN project coordinators later read through the themes to validate that they captured the key views and experiences that the children had expressed within the interviews.

Table 5.6

An Example of a Data Extract from the Interview Transcripts and the Corresponding Codes

Data extract	Coded for
When I didn't use the cards I was really struggling but my mum has said that she noticed I'm really improving since using them.	• Parental involvement • Score increase

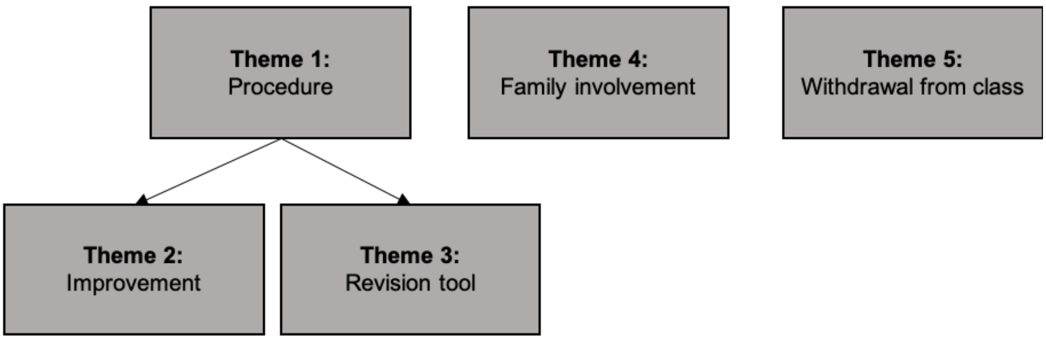

Figure 5.4. Final thematic map showing the seven themes extracted from the interview transcripts.

Results

Theme 1: Procedure

The interviewers asked the children whether they enjoyed used the strategy as a whole, all but three of the children reported that the sessions were fun and enjoyable. That said, the children found certain aspects of the SAFMEDS strategy more positive than others. Positive aspects included the independence the strategy afforded them and the ability to engage in self-paced practice. Children also found the error correction step to be a valuable component of the strategy, as it provided them with a valuable opportunity to develop their skills/knowledge. Engaging with this step enabled the children to answer more cards correctly during consecutive SAFMEDS timings.

> "If you get stuck on a question you can ask the teacher and they'll help you get used to it, so you can develop and improve your scores and get better at it."
>
> "I like doing the cards by myself because I can zoom through all the cards."

With regards to some of the limitations of the strategy, five children reported that with prolonged use the sessions became repetitive and boring. However, most appreciated the need to practice in order to see improvement. At the beginning of each SAFMEDS session, the children or supervisors needed to prepare for the timings (e.g., lay out materials, make sure all of the cards are facing the right way). One child felt like this wasted time within sessions and another felt like they needed a longer session to fit everything in.

> "[My least favourite thing about doing SAFMEDS is] doing things over and over again. Have to in order to get better."
>
> "I think they should put our cards out ready for us when we come. It just wastes time. Some people don't finish 3 goes."
>
> "[We should] come out of lesson to do the cards [rather than tutor period]. Tutor is not enough time [15 minutes]."

Throughout the transcripts there appeared to be conflicting views relating to the 1-minute timing period. Most of the children found the timing period to be an appropriate sprint to enable them to go through the cards. This encouraged them to work through the cards as quickly as possible. Conversely, some children wanted to be able to go through *all* of the cards in the deck and felt that 1-minute was not long enough to achieve this.

> "I like timing it – it's surprising how quickly a minute goes by."
>
> "The people I do it with get fed up because they run out of time."

A group of three children, in Year 4, expressed a preference for tactile flashcards over an unspecified technological alternative. One child in this group liked that the cards provided only one answer to each question; opposed to a multiple-choice selection. For this child, the ability to turn the card over and receive the immediate corrective feedback was a positive aspect of the tactile procedure. Other advantages the children in this group associated with using the physical cards included not having to overcome technological glitches and not having a visible countdown timer.

> "I like that it's cards, because on the computer it gives you options, but on the cards you have to work it out yourself, so if you get it wrong it doesn't matter because you just look on the back of the card."
>
> "If you are on the internet it might keep loading, but with the cards you have to do it at your own pace."
>
> "If you do it on your computer it can pressure you, because on the cards you don't really see the time."

Theme 2: Improvement

Thirty-two of the children we interviewed claimed that using the SAFMEDS strategy helped them to develop their literacy and/or numeracy skills. Across the program, children referenced progressing through different packs, developing confidence, and seeing their scores improve. At a broad level, they felt like this was facilitated by frequent practice and engaging in the error correction procedure. Two children did not feel like they had improved because they felt like they had been practicing skills they had already mastered.

> "[The cards] help me with maths and divide. I used to not know how to divide but now I know a method that counting to the times table of the first one and then you can get the right answers."

> "When I didn't do the cards, I really struggled but now it gets easier and easier."

It is clear from the transcripts that some of the children were motivated by competition during their SAFMEDS sessions. Some children reported being motivated to surpass their personal best scores (self-competition), whilst others were keen to compete with their friends to see who could answer the most cards correctly within 1-minute (peer-competition). Moreover, the transcripts revealed that the children felt a real sense of achievement if they managed to get through all the cards within the deck during a timing; working towards this target helped to keep them motivated. In this context, it may also be important to consider whether children might cheat and write down an inflated score. In order to avoid this someone should supervise each session and children should not publicly announce their scores

> "I kept getting 0 right on the pack but I'm getting better and soon I'll know all the answers."

> "First, I got 5, then 6, then 11. And I beat my partner."

> "We can do it in class and see who has the best score. The competitive element is good."

> "I like counting on how many I got right at the end. I have the motivation to get 60."

> "Some people might get the answer wrong and the teacher isn't looking, so they pretend they got it right, so it is better to have supervision."

> "I would like to keep my score to myself because people might cheat and write your score down on their sheet."

Participation in the XL-LAN program helped children to use their skills outside of the SAFMEDS sessions. Seven children expressed that they were able to apply their new skills (content and fluency) in the classroom, with particular reference to tests and competitions. These children found that they were making fewer mistakes in their class work and had seen their grades start to increase. Three children reported that engaging with the program helped them to develop their confidence; as a result, they felt more able to raise their hand to answer questions in class.

> "If I wasn't doing the cards, I wouldn't be so quick at my times tables."

> "My mental maths gets quicker and my grades are increasing. It makes me feel better about it."

> "I think it has helped my confidence because at the start of Year 4 I wasn't that confident. It makes me confident to answer questions in class."

> "My teacher tells us to do times tables, and I am always sad about it but when I do flashcards, I feel better about it. It was really fun, and I got better at it. And now I know all of my times tables."

> "It has helped me in maths because sometimes we have times tables competitions."

> "My teacher always asks you a question and he asks random people. I'd always be stuck, and I wouldn't get it. When I do flashcards, I always get it now."

Theme 3: Revision tool

Several of the children who participated in the interviews were in secondary education and anticipating exams/tests. They were able to identify the role that SAFMEDS could play in exam revision. The children did not provide further elaboration regarding the skills that they would practice and assess using the SAFMEDS strategy.

> "GCSEs are coming up and will help with revision. Better at revision from cards."

> "It would be useful to use in lessons and classes, helps with revision."

> "I would use them at home to help me revise if I could."

Theme 4: Home use

When the interviewer asked the children if they would use SAFMEDS at home, 28 claimed that they would like to if they were allowed to take the cards home. It became apparent that some children wanted to involve their family in the SAFMEDS process. They felt like their parents could encourage them to use the cards and that they wanted to teach their siblings how to use the strategy so that they could do timings simultaneously. Only five of the children reported that they would not use the cards at home if their provision provided them with the materials to do so. The remaining five children did not respond to this question.

> "I would use them if I could take them home and I think my mum would push me to use them every day."
>
> "I would have a competition with my brother and sister because their good at times tables too."
>
> "I would like to teach my brothers how to use them. If I had 4 packs, I could give them a pack each."

Theme 5: Withdrawal from class

The final theme to emerge from the transcripts encapsulates the idea that some of the children were removed from their lessons to complete their SAFMEDS timings. When removed from classes that they enjoyed, the children showed more resistance and less enthusiasm towards engaging with the SAFMEDS sessions. Conversely, SAFMEDS sessions appeared to be more appealing when the children could leave classes that they perceived to be less important and/or less enjoyable.

> "Sometimes I don't like missing my lesson – it depends on which one it is."
>
> "[I don't enjoy] missing maths lessons, although yesterday was a boring lesson."
>
> "I like missing story time."

Discussion

The children that we interviewed indicated several positive aspects of the SAFMEDS strategy; including independent fast-paced practice, a visual increase in scores, and the opportunity to engage in error correction. Moreover, the children highlighted some points for consideration when rolling out a SAFMEDS program in an applied setting. For example, practitioners may need to consider strategies to reduce cheating, ways to reduce the time it takes to set up each session, and how we can harness an element of novelty when children need to use the strategy for prolonged periods of time.

A growing evidence-base suggests that effective implementation of the SAFMEDS strategy can increase learners' fluency of key academic skills; including numeracy and literacy (e.g., Beverley et al., 2016; Greene et al., 2018; Hunter et al., 2016). During the pilot of the XL-LAN project, children from Year 3 to Year 11 (aged 7-15 years) used the SAFMEDS strategy to improve their fluency of key literacy and numeracy skills. During the interviews they reported that they had developed new strategies to answer questions, had improved their mental recall of facts, and had noticed their progression through SAFMEDS packs. The development of these skills helps to address the cumulative effects of dysfluency by helping children master skills that they did not previously have in their repertoire (Binder, 1996). We anticipate that application of this knowledge can help children access more difficult content on the curriculum.

In the UK, children work towards their GCSE exams in Years 10 and 11, with English language and mathematics forming part of the compulsory curriculum. To the author's knowledge, no published research has reported the use of the SAFMEDS strategy to practice and assess GCSE-level content. However, Hunter et al. (2018) demonstrated that A-Level students (Years 12 and 13) can use the SAFMEDS strategy to support their revision of A-Level psychology facts. The children enrolled on the XL-LAN project highlighted the perceived versatility and applications of the SAFMEDS strategy across ages. This includes the benefits of using the strategy to support class tests and statutory exams. Further quantitative research is necessary to validate the use of the SAFMEDS strategy as a revision tool across all stages of the education system. This would enable researchers to not only explore whether it is consistently effective across different topics (particularly as the content increases in difficulty) but also whether it is as effective as other popular revision strategies.

The effectiveness of a SAFMEDS strategy is not limited to the classroom. Fishel and Ramirez (2005) showed that family involvement is a driver in the success of academic interventions, particularly those supporting literacy and numeracy skills. One advantage of using the SAFMEDS strategy at home is that the person assisting with the timings does not need to know the corresponding answer to the questions on each card (because the learner receives this feedback when they turn the card over). The person supporting the learner with error correction can use a scripted model-lead-test prompting technique if they do not know how to support learning (for example, see Greene et al., 2018). The XL-LAN interviews highlight that some children are keen to use the resources at home and involve their family members in the process. Going forward, this enthusiasm could be utilized to further increase children's skill progression and engagement with the strategy.

When considering progress towards mastery, it is important that children record accurate and honest scores. The majority of existing SAFMEDS literature reports the use of the strategy on a one-on-one basis (i.e., the timings occur between a learner and a supervisor; Quigley, Peterson, Frieder, & Peck, 2017). In larger group designs (e.g., small group or whole class SAFMEDS sessions), it may be more difficult for a supervisor to notice every instance of cheating; particularly when the strategies are subtle or unintentional. Vargas (2013) provided an example of cheating whilst using SAFMEDS during an algebra class. The instructor noticed that some of the highest performing SAFMEDS students were not the highest test performers. When observing a SAFMEDS session, the instructor realized that some of the students were answering the cards too quickly to be attending to all the information provided. It was likely that the students had memorized irrelevant prompts, such as a smudge in the corner of the card or an isolated word. This has clear implications in terms providing children with cards that are too difficult for them and test outcomes. The data from the XL-LAN interviews indicated that children might cheat by placing “not yet” cards in their correct pile or verbally announce a higher score to their peers. This finding is also important in relation to some of the invalid quantitative data that we excluded from the final quantitative analysis; whereby the data was suspected to be invalid either by the supervisor or the XL-LAN coordinators.

The children enrolled on the XL-LAN project indicated that competition motivated them to want to improve their scores. In Hughes, Beverley, and Whitehead’s (2007) study, the instructor encouraged the participating children to aim for their ‘personal best’ score; turning their focus away from peer competition and towards self-competition. By doing this, they hoped that the children would contact the natural occurring reinforcers that occur as a product of learning. It is clear that additional research is needed to explore the impact of peer and self-competition dynamics during a SAFMEDS program. Measures should assess whether peer competition may act as a catalyst for cheating behavior, particularly amongst children who feel embarrassed or disheartened when they achieve a lower score than their peers. Educators may benefit from utilizing individualized target setting to keep children focused on their own progress towards mastery. Aiming to beat their personal best score each session may help limit cheating in a group setting and consequently generate more accurate learning pictures. The more accurate these learning pictures are the easier it will be for educators to make the necessary data-driven decisions that will support the children to become fluent at the skill they are practicing and move onto progressively harder packs (Lindsley, 1995).

Over the course of the XL-LAN program, several children commented on how the SAFMEDS sessions have improved their confidence of numeracy and/or literacy skills. Children's confidence in mathematics and literacy has been strongly correlated with their competence and attainment (Francis et al., 2017; Katzir, Lesaux, & Kim, 2009; Nunes, Bryant, Sylva, & Barros, 2009). The SAFMEDS strategy provides a vessel for learners to practice and assess a skill until they are able to answer questions fluently. Over time, a child can visually see their scores improve towards mastery (Eshleman, 1985). The children we interviewed reported that the SAFMEDS sessions made them feel more confident to answer questions in class. Additionally, they reported that they were making fewer mistakes in their classwork as a result of the fluency-building practice. This further demonstrates some of the transferable benefits of a SAFMEDS program.

General discussion and recommendations

This report aimed to evaluate the effectiveness of the XL-LAN pilot scheme in improving children's fluency of basic literacy and/or numeracy skills. Since its initiation, the XL-LAN project evolved to support 30 schools and organizations. The findings we have presented in this report enabled XLP to strategize about how to further develop the service they provide. Over the course of this pilot, XLP have identified several factors that affect the regularity of SAFMEDS sessions and the fidelity of implementation.

Supervisors, XLP staff, and children highlighted that cheating can occur during a SAFMEDS program. Going forward, the XL-LAN project aims to promote a small group approach to deliver the SAFMEDS sessions. This will enable the supervisors to observe the children more closely and validate their data more regularly. Moreover, XLP will be encouraging supervisors to set the children individualized weekly score targets to help maintain focus and motivating during their SAFMEDS sessions. We anticipate that the focus on self-improvement will reduce unhealthy peer competition and the occurrence of dishonest/invalid scores.

XLP made the decision not to use the SCC to plot data for the duration of the project. They believed that supervisors and children would find it too complex to use. Instead, they opted to use a linear graph—which they felt the audience would be more familiar with. Whilst linear graphs do depict progress, they do not always visualize progress in an accurate and standardized form (Lindsley, 1991c). Learning is more accurately reflected on a multiply/divide scale; whereby we can calculate growth of learning over time (for more information see Calkin, 2005). With the children involved in the XL-LAN project being highly motivated by score increase, it would be worthwhile to upskill provisions to use the SCC during the next academic year to visually display this progress. If a children's SAFMEDS score starts to plateau or worsen, then a change to the instruction, materials, or surrounding environment might be necessary. The data in Study 1 revealed that few children completed more than one SAFMEDS pack during their involvement with the program. Moreover, on average, children's fluency seemed to plateau below a level that would represent mastery. The SCC is a valuable tool that will help supervisors and children identify when they need additional support, and these changes will be instrumental in seeing continuing progress within and across SAFMEDS packs. This will allow children to visualize their learning and also provides the option for educators and learners to set daily goals using trajectory lines (Johnson & Street, 2012).

A final recommendation would be to employ the use of an external standard measure to assess skill performance. This does not aim to replace the use of the SCC, but instead accompany the decisions that provisions make about the skills they choose to focus on. Standard measures such as the Wide Range Achievement Test, Fourth Edition (WRAT-4;

Wilkinson & Robertson, 2006) offer administrators a way to assess and monitor literacy and numeracy skills. The advantage of these tests is that the outputs provide details about where children's skill deficits lie, comparative to their age. By collecting data at the beginning of the project, XLP will be able to support provisions in their choice of SAFMEDS packs. The sub-scales of these assessments can be re-administered whenever necessary; providing further indication of progress and ensuring that the children are able to apply the skills that they have learnt from the SAFMEDS cards to another context.

XLP also offer a youth bus service that hosts computer equipment for homework and aims to provide a comfortable space for young people to relax in (XLP, n.d.c). Over the course of the XL-LAN project, the delivery of SAFMEDS strategy on board this service has evolved to try and engage children and encourage them to come back and practice skills. Most recently, XLP have employed a 'loyalty card' system, whereby young people receive a stamp each time they come to the bus to engage with the SAFMEDS strategy. This acts as a score card, so the children can see their tabulated progress. The data for this project was not included in the final analysis due to the impractical nature of gaining consent. Nevertheless, this is a novel approach to encourage young people to engage with the SAFMEDS strategy.

The relationship between XLP and Bangor University has been reciprocal throughout this project. The XL-LAN project has provided several valuable outputs, not only in relation to the skill development of the children but also directions for future research and the evolution of the literacy and numeracy support they continue to offer to educational provisions. This work highlights the benefits associated with charities and education provisions working with researchers to develop services.

Chapter 6 : **Using an instructional fluency approach to teach addition skills in a pupil referral unit: A pilot study**

Owen, K., Watkins, R. C., Beverley, M., & Hughes, J. C. (accepted/in press). Using an instructional fluency approach to teach addition skills in a pupil referral unit: A pilot study. *Wales Journal of Education, 22*(2).

Preface

Figure 2.4 outlines some of the context and rationale for the research that we present in this chapter. Estyn (2017) revealed that four in ten PRUs in Wales did not sufficiently support children to develop the numeracy skills they need to support their learning across the curriculum. With a high proportion of children arriving at PRUs with additional and complex learning needs (Department for Children, Schools and Families, 2008; Estyn, 2015; Welsh Government, 2018), is important to identify effective interventions that can help them to catch up with their mainstream peers and leave education with necessary numeracy skills. The first author was interested in piloting the use of DI and PT to help support and accelerate children's learning of arithmetic in PRUs. The *Corrective Mathematics* (CM) program offers a way of delivering scaffolded and unambiguous lessons to help children master content. To complement this, PT time trials enable learners to practice these skills to fluency and for a practitioner to make data-driven decisions to support progress. This is the first known attempt to pilot these methods in a PRU to support mathematics development.

This chapter contains the manuscript of the paper that we have submitted for publication. At the time of the submission of this thesis, this paper has undergone peer review and it is in press. We have edited the version that we have presented here to adhere with APA formatting guidelines (including placing figures and tables in the main body of the text). In Chapter 7 we discuss the implications, applications, and limitations of this research in more detail.

Abstract

Pupil referral units (PRUs) in Wales accommodate children who present with a range of difficulties that cannot be managed within a mainstream setting. Many children attending PRUs in Wales do not develop the numeracy skills that they need to support their learning across the curriculum. In an effort to teach and assess addition skills, the authors assessed the effects of using a combination of direct instruction (DI) and precision teaching (PT) in a PRU. Over six school weeks, we worked with five children (aged 7 to 10 years) on a one-on-one basis through the *Corrective Mathematics* addition curriculum (Engelmann & Carnine, 2005). Following each lesson, the children completed an individualized fluency assessment, which we tailored to their needs using PT methods. We collected baseline and follow-up data using the Test of Early Mathematics Ability (TEMA-3), the Wide Range Achievement Test (WRAT-4) and the *Corrective Mathematics* placement test. We also interviewed the children post-intervention to gain insight into their experience of the approach. The results provide evidence to support the use of an instructional fluency approach in a PRU setting to help children develop early mathematics skills, particularly for children who engaged in the sessions regularly. Due to the small sample size, the results of this study have limited generalizability but may help shape future research investigating effective strategies for teaching mathematics in PRUs.

Pupil referral units (PRUs) accommodate children with complex needs that cannot be managed within mainstream school settings. Children attend a PRU for a variety of reasons relating to behaviour or illness; with a large proportion holding a diagnosis for an additional learning need (Estyn, 2015; Welsh Government, 2018). In their report, Estyn (2017) indicated that there are currently 25 registered PRUs in Wales, educating approximately 665 children. Estyn judged the quality of provision to be adequate in 36% of these PRUs and unsatisfactory in a further 14%. This was mainly due to the wide variability of outcomes children achieve (including the narrow breadth of qualifications) and low levels of participation and engagement in learning. Their inspection report also revealed that four in ten PRUs taught numeracy skills targeted towards early development and that many of the children did not develop the skills they needed to support their learning across the curriculum. As a result, the curriculum these units offered lacked challenge, did not ensure that all children in the classroom achieved, and that the pace of learning was too slow.

The term *mathematics* refers to an international discipline, which integrates concepts, rules and procedures involving quantities and symbols. The term *numeracy* refers to the application of mathematical knowledge to every-day life (Resnick & Ford, 1981/2008; Welsh Government, 2019). Longitudinal studies provide evidence to suggest that that acquisition of age-expected mathematics and numeracy skills during early childhood is predictive of later mathematical achievement (Desoete, Ceulemans, De Weerdt, & Pieters, 2010; Missall, Mercer, Martínez, & Casebeer, 2012). Underachievement in mathematics and numeracy also has wider implications with regards to access to further education and employment prospects (Banerjee, 2016; Geary, 2011). This highlights the need to identify pedagogical approaches that can help improve the mathematics outcomes of children who attend PRUs.

The Foundation Phase Framework (Welsh Government, 2015) and the *Mathematics Programme of Study* in Wales (Welsh Government, 2016) describe the outcomes that children should be able to achieve at the end of each school year in Wales. By the end of foundation phase (Year 2, aged 7 years), children should be able to solve simple addition problems. Examples of these skills include being able to solve one-step word problems that involve addition (e.g., 7 + __ = 9); using known number facts when adding three single digit numbers; and being able to mentally add 10 to a given number up to 100 (Welsh Government, 2015). By the end of primary education (Year 6, aged 11 years), the curriculum prescribes that children should be able to identify missing numbers a sequence using their addition skills and simplify formulae involving the addition of variables (Welsh Government, 2016).

The new Curriculum for Wales (Donaldson, 2015) places an emphasis on making mathematics and numeracy experiences engaging and accessible for all children. Schools in Wales have the statutory duty to teach numeracy across the curriculum to help children to apply their skills and solve problems in real-world contexts (Welsh Government, 2013a). The Curriculum for Wales 2022 guidance identifies that children need to be able to fluently use the four basic arithmetic operations (addition, subtraction, multiplication, and division) and understand the relationship between them (Welsh Government, 2019). Jordan, Hanich, and Kaplan (2003) demonstrated that children aged 7 to 9 years who were unable to perform arithmetic facts fluently (i.e., reach the correct answer in less than 3-seconds) performed significantly lower on mathematics tests compared to age-matched children who were more fluent. These tests included questions encompassing story problems, place value, forced retrieval of number facts, calculation principles, and written computation. The Education Endowment Foundation (2017) recommend that schools support children to develop fluent recall of mathematics facts pertaining to the four arithmetic operations. They acknowledged that without this knowledge children are likely to encounter difficulty understanding and using the mathematical concepts taught later in the curriculum.

Binder, Haughton, and Bateman (2002) explained that schools traditionally view children achieving 100% accuracy as “mastery”, but with additional practice children can recall facts both quickly and accurately (i.e., fluently). Without the ability to perform basic skills fluently children are likely to struggle to master complex skills (Johnson & Layng, 1996; Nelson, Burns, Kanive, & Ysseldyke, 2013). For example, if a child is not able to recognize numbers fluently, they will be unable to read single digit addition sums without hesitation. Binder (1996) coined this phenomenon *cumulative dysfluency* and believed it can explain academic underachievement and failure within education. There are several evidence-based strategies that aim to remediate cumulative dysfluency; including direct instruction (DI; Kozioff, LaNunziata, Cowardin, & Bessellieu, 2000) and precision teaching (PT; Gallagher, Bones, & Lombe, 2006).

The importance of teachers considering a range of teaching approaches, including more direct teaching, is recognized as one of the 12 pedagogical principles in the Curriculum for Wales (Donaldson, 2015). DI is an approach which aims to teach component skills to mastery. Using a combination of behavioral techniques (e.g., reinforcement principles and task analysis), DI programs teach learners skills in a sequential, explicit, and scaffolded order (Kinder & Carnine, 1991). DI lessons incorporate teacher demonstrations and guided practice to establish learner independence and skill acquisition (Archer & Hughes, 2011). Several studies have demonstrated that commercially available DI programs can help children learn basic mathematics skills; including *Corrective Mathematics* (Glang, Singer, Colley, & Tish, 1991; Parsons, Marchand-Martella, Waldron-Soler, Martella, & Lignugaris-Kraft, 2004; Sommers, 1991). These programs aim to supplement classroom teaching rather than replacing it by focusing on children’s skill deficits. This makes DI programs useful for both mainstream learners who have fallen behind age-expected norms and those with additional learning needs (Flores & Kaylor, 2007). In their meta-analysis, Stockard, Wood, Coughlin, and Khoury (2018) demonstrated that DI programs can significantly improve academic outcomes across the curriculum. Analysis of 328 studies across a 60-year period revealed that DI approaches consistently yield positive effect sizes, with most estimates falling within the range of medium or large. This provides strong evidence to support the use of DI approaches in education, including its use to teach component mathematics skills to children.

A complementary approach to support skill development is precision teaching (PT). PT is a method of assessment that aids decisions about subsequent instruction. There is also a specific focus on building fluency to help children master skills to a level that promotes maintenance, endurance, stability, application, and generalizability (Johnson & Layng, 1992; Johnson & Street, 2012). Within a PT approach, educational practitioners identify the skill(s) they want to help their learner develop, they provide opportunities to practice the skills, record data, and try new/altered techniques to teach skills when necessary (White, 1986). These decisions are guided by learning pictures that emerge on a standard celeration chart (SCC; see Lindsley, 1995). For example, if a child's data demonstrates that they are answering less questions correctly within 1-minute over several consecutive days, then a practitioner should consider changing the task they have set. A feasible suggestion might be to assess if the child has mastered all of the associated prerequisite skills (Kerr, Smyth, & McDowell, 2003). Chiesa and Robertson (2000) demonstrated that employing PT methods can support children's mathematical fluency development. The results from their study suggest that PT driven fluency training (including daily practice, time probes, and individually tailored materials) can help children to rapidly improve their mathematics outcomes.

Both PT and DI approaches focus on behavioral mastery and fluency. Desjardins and Slocum (1993) argued that integrating PT methods into DI programs can help learners to establish mastery of key concepts. Combing these instructional technologies enables learners to fluently perform basic skills, provides efficient and effective practice opportunities, and ensures that learners are able to perform skills at a proficient level before progressing onto more difficult ones. An emerging evidence-base reports the benefits of using an instructional fluency approach to teach literacy skills to mastery (e.g., Adda Ragnarsdóttir, 2007; Hulson-Jones, Hughes, Hastings, & Beverley, 2013). The available data in this field suggest that the approach can support a variety of learners, including children who attend mainstream primary schools (Kubina, Commons, & Heckard, 2009) and children with additional learning needs (Morrell, Morrell, & Kubina, 1995) to develop fundamental literacy skills. However, no known published research has investigated whether this approach can be applied to a manualized DI mathematics curriculum and elicit positive outcomes for children who attend a PRU.

The current small-scale exploratory study aimed to assess the effects of using an instructional fluency approach to teach and build fluency of addition skills. The authors used a commercially available DI program—*Corrective Mathematics* (CM)—in a PRU situated in North Wales. At the end of each CM lesson, the children completed a 1-minute timing to assess their fluency of basic addition skills (adhering to PT methods). We assessed the children’s numeracy gains over six weeks of instructional fluency sessions. This paper also includes data relating to session attendance, the children’s literacy skills, as well as the children’s attitudes towards the instructional fluency intervention.

Method

Ethics

This study received full ethical approval from the School of Psychology’s research ethics committee at Bangor University (reference number: 2018-16417). After approaching a PRU in North Wales to take part in this research, we sent opt-in consent forms to the headteacher and children’s parents/guardians. Children provided assent to complete the assessments and intervention sessions. If at any point they did not want to participate they returned to their classroom. We later gave children the opportunity to return and complete the given task. Throughout this paper we refer to each child by a pseudonym to protect their anonymity.

Sample

We received parental consent to assess ten of the children who attended the PRU. Following the baseline assessments, we identified five children (aged 7 to 10 years) to participate in the instructional fluency intervention. We selected these children on the basis that they attended the unit on the days that the first author was able to conduct the intervention sessions and they completed all baseline measures. This allowed us to ensure that the children had the appropriate prerequisite skills to participate in the intervention (e.g., they could recognize numbers, write numbers independently, and were able to read basic addition problems independently). The assessment also enabled us to identify the children who would benefit from the CM addition program—these children all placed on a lesson within the addition module. Table 6.1 outlines the characteristics of the children who participated in the intervention.

Table 6.1

Characteristics of the Children who Participated in the Instructional Fluency Intervention at Baseline

Child	Diagnostic label	Chronological age
Tom	ASD, ADHD	7-years 8-months
Dean	ADHD	10-years 1-month
Will	-	10-years 1-month
Chris	ASD, ADHD	9-years 10-months
Leo	-	9-years 11-months

Note. ASD = Autism Spectrum Disorder, ADHD = Attention Deficit Hyperactivity Disorder

Assessments

The researchers who administered the assessments were blind to the aims of the project and were not involved in delivering the intervention sessions; the purpose of this was to minimize administration bias. The children completed all of the assessments on a one-on-one basis with a researcher in a quiet room in the PRU they attended. In order to reduce the effects of fatigue and the likelihood of behaviours that challenge occurring as a result of demand, the researchers ensured that the children had sufficient breaks between assessments and sub-tests. None of the children completed all of the assessments in one sitting but did complete them within a 1-week period. Following the completion of each assessment, the researchers rewarded each child with verbal praise and a sticker. The baseline assessments for this study took place in April 2019. Following 7 weeks (inclusive of 6 weeks of intervention sessions and 1 week of half term), the children completed the follow-up assessments in June 2019.

Test of Early Mathematics Ability (TEMA-3)

The TEMA-3 (Ginsburg & Baroody, 2003) identifies children who are likely to develop numeracy difficulties. Typically used with children aged between 3-years 0-months and 8-years 11-months, this assessment offers insight into children's ability to perform the mathematics skills that are typically taught during early schooling (e.g., reading numbers, counting forwards and backwards, using finger displays, and using a number line).

The TEMA-3 offers an entry point for the assessment based on the child's age. We used this recommendation to limit administration time. A researcher worked forward through the test items from the age entry point until the child reached a ceiling (i.e., they answered five consecutive items incorrectly). The researcher also ensured that they had a basal measurement for each child (i.e., they had answered at least five consecutive items correctly);

in some cases, this required the researcher and child to work backwards from the age entry point.

The researcher sat opposite the child across a table with the picture book and examiner record booklet. Each item on the TEMA-3 has a script for assessors to follow. The researcher read this out loud and waited for the children to respond. Some of the questions required the children to respond using their fingers, answer orally, or provide a written response.

To account for repeated administration, The TEMA-3 offers two parallel test forms. Bliss (2006) reported that these forms have high levels of internal reliability (α); with previous research reporting reliability coefficients between .92 and .96. At baseline the children completed Form A and at follow-up they completed Form B.

Wide Range Achievement Test (WRAT-4)

The WRAT-4 (Wilkinson & Robertson, 2006) provides a battery of measures assessing reading, sentence comprehension, and mathematical computation. The WRAT-4 assesses an individual's ability to decode letters and words, gain meaning from words, count, identify numbers, solve oral mathematics problems, and calculate written mathematics problems (from basic arithmetic to advanced operations). This is a norm-referenced assessment which practitioners can use with individuals aged 5-years through to 94-years.

For this study, a researcher worked through the script that accompanies the administration of each sub-test (reading, comprehension, and mathematics). The children responded either verbally or in written form on the corresponding test form. Some of the children were unable to read passages of text independently. The researcher read the comprehension passages and literacy-based mathematics questions out loud for these children.

The WRAT-4 offers two parallel test forms for repeated administration. Dell, Harrold and Dell (2008) indicated that the forms have high internal consistency; with reliability coefficients ranging from .92 to .98. Within this study, the children completed the Blue form at baseline and the Green form at follow-up.

Corrective Mathematics placement tests

The CM program is comprised of systematically sequenced lessons for key mathematics skills. In order to place children on a lesson that meets their needs, they can sit the CM placement test. This test is a paper-based assessment and requires children to write their responses on the test form. A researcher administered the CM placement test to identify if all of the children met the criteria for the addition module (i.e., they made more than 1 error on Part A of the assessment). The data from this assessment also enabled us to place the children on an appropriate lesson within the CM addition program.

All of the children completed Parts A and B of the assessment. Part A assessed the children's ability to answer addition sums in columns; starting with single to single digit addition and progressing onto adding four multi-digit numbers together. Part B assessed the children's ability to answer subtraction calculations (including single to single digit and double to double digit calculations), as well as their ability to answer subtraction word problems. A researcher read the word questions to the children if they were unable to read independently.

The children had 20-minutes to answer as many questions collectively from Part A and B as they could. If they identified that they could not answer anymore of the questions before the end of the timing period, the researcher stopped the assessment and scored their responses. If any of children made one error, or less, on both Parts A and B, then they would have met the criteria to progress onto Parts C (multiplication) and D (division); however, none of the children met this threshold at baseline or follow-up.

Follow-up interviews

Following the intervention, the first author interviewed each of the children who participated in the instructional fluency intervention. The informal interviews also took place on a one-on-one basis in a quiet room within the PRU. The first author asked the children if they had enjoyed taking part in the intervention, what aspects about it they liked or disliked, and if they would like to continue taking part in the intervention sessions in the future. When appropriate, the first author asked the children to elaborate further on their answers and/or asked follow-up questions. Please refer to Appendix E for a list of the predetermined questions and prompts.

Materials

DI program

Corrective Mathematics (CM; Engelmann & Carnine, 2005) is a commercially available DI program that offers seven modules to build children's understanding of key mathematics skills. For this study, we focused on the addition module and aimed to complete one lesson per session. During each lesson, the first author used the presentation book, which contained a script for each exercise and for correcting children's errors. The exercises within each lesson required the children to respond either verbally, by pointing to an answer, or by writing the answer down. The lesson-specific worksheets provided the children with an opportunity to practice and review their skills throughout the program. The children came out of class to complete the sessions in a separate room within the PRU.

Randomized practice sheets

To support the PT element of the sessions, the first author generated a collection of addition practice sheets. Each sheet contained 30 random column addition sums, tailored to the ability of each child (e.g., all single digit addition combinations containing digits 0 to 9; exclusively +0 and +1 sums). In order to complete the worksheet, the children had to write the correct answer underneath each column sum. They had 1-minute to answer as many questions as they could; working from left to right. They could skip questions if they did not know the answers fluently and return to them at the end if time permitted them to do so. The children completed one practice sheet following each CM lesson. The first author scored the answers based on the number of correct digits written within 1-minute.

The Standard Celeration Chart (SCC)

The first author plotted the children's scores from the randomized practice sheet activity onto their individualized SCCs. This enabled the first author to make decisions about whether each child was making sufficient progress across sessions or if the activity needed to be altered. As a general rule, if a child did not make desired progress (i.e., their score decreased or maintained) over three consecutive sessions, the first author made a change to the practice sheet activity (e.g., altered the content of the worksheets or provided the children with some further instruction to help them answer specific questions).

Results

Attendance

The children had the opportunity to attend three intervention sessions a week, for six school weeks. Tom, Will, and Leo started but did not complete one of their sessions due to refusal to comply. In these instances, the first author terminated the lesson and the children returned to their classroom. Tom and Will repeated the CM lesson that they did not complete in the following session. Despite given the opportunity, Leo refused to attend any more sessions for the remainder of the intervention period. Table 6.2 displays the total number of sessions each child attended.

There are several reasons why some of the children did not attend all of the sessions. Reasons for non-attendance included: refusal to leave the classroom (i.e., lack of assent), school trip, illness, or a competing activity within school that required the child's participation (e.g., another intervention). Some of the children also attended a mainstream school for half a day throughout the week, so they were not always in the unit to attend the sessions due to timetabling changes.

Table 6.2

Progress Through the CM Curriculum Over the Intervention Period

Child	Starting CM lesson	End CM lesson	Total number of sessions attended
Tom	1	9	9.5
Dean	1	14	14
Will	1	16	16.5
Chris	1	11	11
Leo	23	26	3.5

Note. Some of the children started a lesson but did not complete them due to refusal; this is denoted by 0.5.

TEMA-3

The raw scores from the TEMA-3 assessment can be found in Table 6.3. A Wilcoxon signed-rank test revealed that the children's TEMA-3 raw scores did not vary significantly following the instructional fluency intervention ($Z = 1.63$, $p = .10$). It is worth noting that this analysis does not account for the variance in session attendance across the five children. However, the age equivalence data demonstrated that all of the children who attended the instructional fluency sessions, except Leo, made greater gains than would be typically expected over a 7-week period. Leo did not improve on this measure, but also engaged in the fewest number of sessions (completing only three CM lessons across 2-weeks).

Figure 6.1 shows evidence of Tom's development between baseline and follow-up. Despite reversing the numbers in his answer, his single digit addition computation skills improved. Although he did not demonstrate digit reversal during the intervention sessions, he made this mistake consistently across the TEMA-3 follow-up assessment. This affected his overall raw score performance on the TEMA-3 and is reflected in his age-equivalence outcomes.

Table 6.3

The Children's Baseline and Follow-Up Outcomes on the TEMA-3.

	Raw score			Age equivalence		
Child	Baseline	Follow-up	Difference	Baseline	Follow-up	Difference
Tom	47	52	5	7- years 0-months	7-years 9-months	9-months
Dean	49	53	4	7-years 3-months	7-years 9-months	6-months
Will	33	48	15	6-years 0-months	7-years 3-months	1-year 3-months
Chris	45	46	1	6-years 9-months	7-years 0-months	3-months
Leo	64	63	-1	8-years 9-months	8-years 9-months	0-months

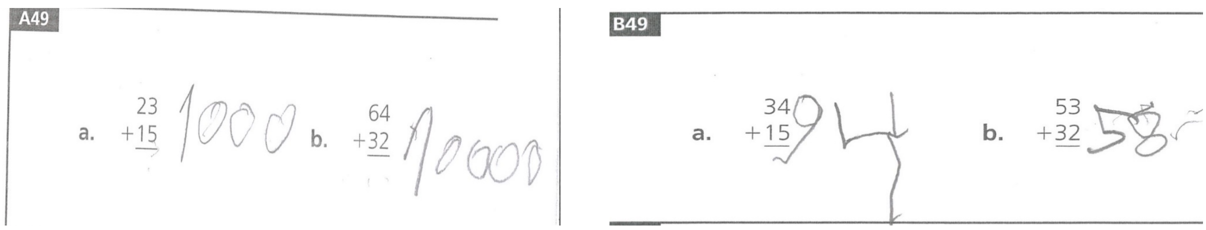

Figure 6.1. Tom's responses on the TEMA-3 at baseline (left) and follow-up (right)

WRAT-4

A Wilcoxon signed-rank test revealed that the children's standard scores did not significantly improve on the mathematics sub-measure between baseline and follow-up (Z = 0.41, p = .69). Whilst the CM program aims to help children master basic mathematics concepts, there is a level of literacy involved in reading and solving mathematics problems (e.g., reading word problems and deducing the corresponding calculation). Therefore, it is important to consider literacy skills in the wider context of numeracy development. We conducted signed-rank test on the children's standard reading (Z = 0.00, p = 1.00) and comprehension scores (Z = -0.27, p = .79). Neither analysis revealed a significant improvement between baseline and follow-up. Table 6.4 outlines the children's standard scores on the mathematics and literacy sub-tests.

Table 6.4

The Children's Baseline and Follow-Up Standard Scores on the WRAT-4

Child	**Word reading**			**Comprehension**			**Mathematics**		
	Baseline [95% CI]	Follow-up [95% CI]	Difference	Baseline [95% CI]	Follow-up [95% CI]	Difference	Baseline [95% CI]	Follow-up [95% CI]	Difference
Tom	64 [57, 74]	76 [69, 85]	12	107 [99, 115]	107 [101, 112]	0	88 [79, 98]	83 [74, 95]	-5
Dean	79 [72, 88]	80 [73, 89]	9	84 [77, 92]	78 [73, 84]	-6	67 [59, 79]	65 [55, 76]	-2
Will	68 [61, 78]	68 [61, 78]	0	96 [88, 104]	97 [91, 103]	1	61 [53, 73]	68 [61, 78]	7
Chris	98 [90, 106]	96 [89, 103]	-2	100 [92, 108]	87 [82, 93]	-13	56 [49, 68]	69 [60, 82]	13
Leo	94 [86, 103]	89 [81, 98]	-5	102 [94, 110]	111 [105, 116]	9	87 [78, 97]	84 [74, 96]	-3

CM placement test

The scoring system for the CM program considers the number of errors children make whilst solving computation problems. Part A focuses on addition skills. A Wilcoxon signed-rank test revealed that the children's scores Part A did not significantly vary across the intervention period ($Z = -0.16$, $p = .88$). Part B assesses the children's ability to answer subtraction calculations. Our results suggest that as a group, the children did not significantly reduce the number of errors they made on Part B between baseline and follow-up ($Z = -0.85$, $p = .40$). Table 6.5 outlines the children's individual progress across Parts A and B.

Table 6.5

The Number of Errors Each Child Made on the CM Placement Tests

	Part A			Part B			Addition Lesson placement
Child	Baseline	Follow-up	Difference	Baseline	Follow-up	Difference	Baseline
Tom	8	8	0	11	4	-7	1
Dean	8	8	0	9	10	1	1
Will	8	4	-4	10	10	0	1
Chris	8	8	0	13	10	-3	1
Leo	3	4	1	9	9	0	23

SCC data

Here we present Dean's SCC as an example to illustrate its use by the first author across the intervention period (see Figure 6.2). Refer to Appendix F to see Tom's, Will's, Leo's, and Chris' fluency progress.

Dean made limited progress on the single digit addition practice sheets over the first two weeks (celeration: x1.05, bounce: x1.3). It appeared that he was struggling to answer questions where the answer exceeded 10 and he needed to use his fingers to count. The first author altered the practice sheet activity to focus on building fluency on single digit addition sums where the answer did not exceed ten first, with the plan to reintegrate more difficult sums following their introduction in the CM program. This alteration saw an improvement in

Dean’s progress (celeration: x1.1). After showing limited acceleration in his correct responses after half term, the first author made the decision to simplify the activity even further; by focusing on adding 0, 1, and 2 to numbers 0 through 9. This saw an initial improvement in Dean’s data before the intervention ended.

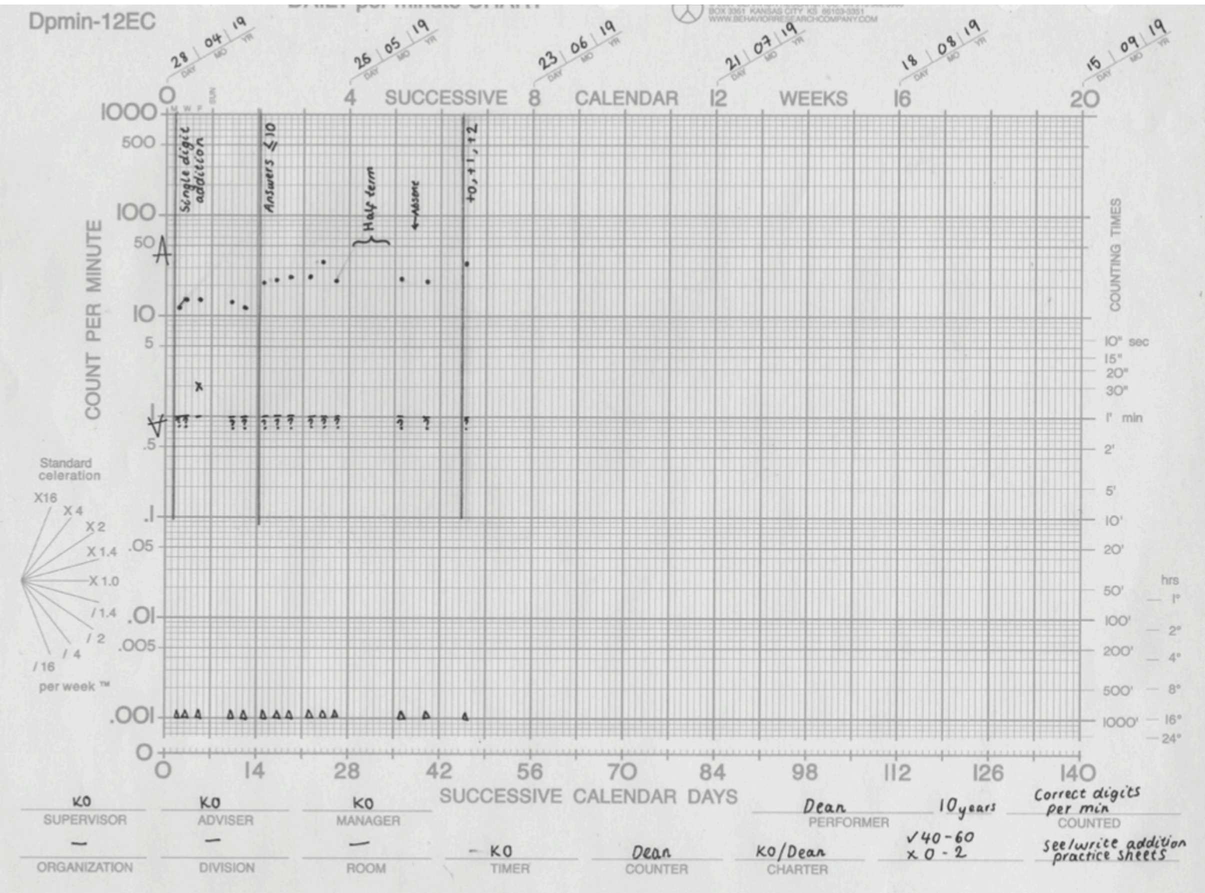

Figure 6.2. Dean's SCC. The dots represent the number of correct digits Dean wrote in 1-minute, the crosses refer to the number of incorrect digits in 1-minute (with the question marks denoting zero errors), and the triangles depict the number of timings Dean completed each session (i.e., one timing a day).

Follow-up interviews

When asked which aspects of the intervention they enjoyed, Tom reported that he liked seeing his progress reflected on the SCC (particularly in reference to the correct responses going up), doing the work well, and receiving stickers for taking part. Will found the sessions fun and enjoyed learning new things and Leo enjoyed the mathematics content that we covered within the sessions.

With regards to the elements of the session they enjoyed the least, Dean expressed that he did not like the 1-minute fluency timings we completed after each lesson. Despite not enjoying this aspect, he acknowledged that it was a useful element of the session as he was now able to write down numbers faster. Chris claimed that he did not enjoy the sessions due to the repetitiveness of the content. Will did not enjoy answering sums containing big numbers.

All five of the children appreciated that the intervention was useful for them. Dean, Will, and Tom claimed that the content they had learnt and practiced within the sessions had helped them with their classwork. Will and Tom felt like taking part in the intervention helped them to get smarter. Dean and Chris both identified that the sessions were useful in the sense that they helped them to learn.

Dean, Will, and Tom indicated that they would like to carry on using the instructional fluency approach to help them learn mathematics. Leo and Chris did not want to engage with the program anymore; with Leo indicating that he felt like the sessions took too long. Leo and Tom both revealed that they did not always engage with the sessions due to a more appealing activity being available in their classroom (i.e., they chose to play a game with their peers rather than completing an instructional fluency session).

Discussion

Estyn (2017) identified that half of the PRUs in Wales are not supporting the children they accommodate to satisfactory standards. Many of the children who attend these units are unable to perform age-expected mathematics skills, they display a lack of interest towards learning, and the pace of learning is too slow for them. Through this small-scale exploratory study, the authors aimed to investigate whether children attending a PRU would engage with an instructional fluency intervention targeting addition skills and whether it would help accelerate their mathematics outcomes. To the authors' knowledge this is the first study to investigate the use of this combined instructional technology to teach mathematics skills a in PRU.

The TEMA-3 results suggest that a regular attendance to instructional fluency sessions can help children attending a PRU to learn some of the mathematics skills that they would typically be expected to have learnt and acquired during early childhood. However, the approach did not significantly affect the children's standard scores on the WRAT-4 assessment. This may suggest that six-weeks of the intervention is not sufficient to help children access higher-level mathematics content or develop comprehension skills. Three of the five children who participated in the intervention indicated that they would like to continue using the instructional fluency approach to learn mathematics. Overall, the results from this study suggest that it is feasible to use an instructional fluency approach on a one-on-one basis in a PRU to improve mathematics outcomes. The outcomes appear to be contingent on the children complying to attend and showing willingness to engage with the intervention sessions.

DI programs aim to identify and teach to children's skill deficits, using explicit and scaffolded teaching methods. That is, DI programs teach skills in a sequential order based on the premise that children have to master certain prerequisite skills in order to access higher-order content (Kinder & Carnine, 1991). Previous research has demonstrated that DI is an effective remedial intervention for children in mainstream schools and children with additional learning needs (Flores & Kaylor, 2007). After six weeks of instructional fluency sessions, Will, Tom, Dean, and Chris increased their age equivalence scores on the TEMA-3 by at least 3-months. This suggests that the intervention helped the children attending the PRU to master some of the mathematics skills that children typically acquire during early schooling. This data supports the contention that using an instructional fluency approach can accelerate learning and remediate skill deficits (see, for example, Morrell, Morrell, & Kubina, 1999). Moreover, this finding is in line with the wider DI and PT literature which provides strong evidence to support the benefits of using these technologies in education (see, for example, Stockard et al., 2018; Chiesa & Robertson, 2000).

Administering the WRAT-4 enabled us to assess the children’s literacy and numeracy abilities comparative to age-expected norms. The data suggest that all of the children who participated in the intervention had low-level word reading skills for their age at baseline; with Dean and Will also displaying low-level comprehension skills. This may have wider implications with regards to the children’s ability to access specific mathematics content on the curriculum. Passolunghi and Pazzagila (2005) suggested that solving reading comprehension questions and mathematics word problems require children to use the same problem-solving skills. In order to answer these types of questions, a child’s working memory needs to process the relevant text and ignore irrelevant information. This might explain why the children who participated in this study were unable to answer the mathematics word problems on both the WRAT-4 and CM placement test (Part B). The CM addition program integrates word problems and associated comprehension strategies from lesson 19. We only ran this intervention for six school weeks, meaning Tom, Dean, Will, and Chris did not engage with these lessons within the CM addition curriculum. Future research could extend the intervention period and investigate the effects of the lessons on children’s literacy and numeracy skills.

Estyn (2015) noted that a large proportion of children who attend PRUs in Wales have additional learning needs. In our sample, Tom, Chris, and Dean held a diagnostic label for autism spectrum disorder (ASD) and/or attention deficit disorder (ADHD). Children with underlying developmental disorders often display attention difficulties which can make them more suspectable to poor academic outcomes and long-term behavioral problems (May, Rinehart, Wilding, & Cornish, 2013). Jordan and Levine (2009) identified five mathematical competencies that children typically acquire during early childhood: the ability to rapidly recall small qualities up to four items; counting abilities; magnitude comparison; estimation, and; arithmetic operations. These early numerical competencies provide the foundation for later mathematics skills to be built upon (Geary, 2000). Titeca, Roeyers, Josephy, Ceulemans, and Desoete (2014) identified that preschool children with high-functioning ASD perform at the same level as typically developing children on these numerical competencies. When it comes to the higher-order mathematics skills, children aged 6 to 7 years old with ASD perform significantly lower than their typically developing peers on questions pertaining to number fact retrieval and word problems. This theory might explain some of the disparity between the children's chronological age and age-equivalence on the TEMA-3 assessment within the current study. The items on the TEMA-3 assessment increase in complexity from the age entry point, so without mastery of the foundation numerical competencies (e.g., counting objects) it is possible that the children who participated in our study were unable to tackle the questions that integrate the higher-order skills despite their chronological age.

All of the children who participated in this study also performed below average for their age (i.e., a standard score < 100) on the WRAT-4 mathematics sub-test. Without sufficient mastery of early mathematical skills, the children who participated in this study might have been unable to understand the concepts and procedures underlying more complex mathematics problems. Both DI (Celik & Vuran, 2014; Rockwell, Griffin, & Jones, 2011; Thompson, Wood, Test, & Cease-Cook, 2012) and PT (Brady & Kubina, 2010) have documented benefits when researchers have used the strategies with children with developmental disabilities, including ASD and ADHD. Limited research in this field has demonstrated that educators can use DI and PT in conjunction with one another to help remediate mathematical skill deficits amongst populations with additional learning needs (see, for example, Delli Sante, McLaughlin, & Weber, 2001).

Of the five children who participated in the intervention, three indicated that they would like to continue using the approach to improve their mathematics skills. This finding suggests that the children associated some level of social validity with the intervention. Extensions of this research should consider collecting further data to assess common aspects of the intervention that children enjoy and elements that might need further refinement. Researchers could use these data to develop strategies to enhance children's engagement with instructional fluency sessions. This extension may also help identify some of the barriers in education that prevent children attending PRUs from engaging in similar intervention programs (e.g., competing classroom activities).

Some of the children who participated in this study attended the PRU on a part-time basis; spending a percentage of their time in mainstream primary school. It was not possible to gather information on the strategies used to teach mathematics in each school and compare these to the instructional fluency approach described in this paper. We appreciate that the findings and conclusions drawn from the current study would have been enhanced by the inclusion of a control group. The inclusion of a control group would have helped established whether the instructional fluency approach has any additional advantages for children attending PRUs, compared to the typical classroom teaching that they are exposed to. Due to the variability in the children's ages, diagnostic labels, the percentage of time they attend the PRU, and their baseline mathematical abilities, we were unable to match the children who returned their parental consent forms. In addition to this, we had a small dataset as a result of some of the children being unable to complete all of the necessary assessments; this was due to lack of assent and/or non-attendance. Due to the time constraints surrounding this project we were also unable to explore the possibility of the children acting as their own controls and measuring their performance growth before and after exposure to the intervention. Future replications of this research should consider the recruitment of control data to establish the differences in performance gains between children in PRUs who attend instructional fluency sessions regularly compared to the effects of their typical exposure to mathematics instruction.

The researchers who conducted the assessments for this study were unaware of the aims of this project and which children had been selected to participate in the intervention. We hoped that this would reduce any confounds surrounding administration bias. However, it is important to consider that the children might have altered their behavior as a result of their participation in this study. McCarney et al. (2007) explained that it is important for researchers to consider the impact of the Hawthorne effect in relation to the generalizability of research to day-to-day life. In the context of the current study, it is possible the children's performance is an underrepresentation or overrepresentation of their performance in the classroom due to their reactivity to the testing conditions.

The data presented in this paper suggest that the instructional fluency approach can support children in a PRU to improve their basic mathematics skills. There is a strong evidence-base supporting the use of DI to teach children literacy (Przychodzin-Havis et al, 2005; Simonsen & Gunter, 2001) and numeracy skills (Przychodzin, Marchand-Martella, Martella, & Azim, 2004). Moreover, practitioners have used PT methods to record and monitor performance for many academic skills, such as mathematics (Chiesa & Robertson, 2000), reading (Hughes, Beverley, & Whitehead, 2007), content-specific terminology (Beverley, Hughes, & Hastings, 2009; Stockwell & Eshelman, 2010), and second language acquisition (Beverley, Hughes, & Hastings, 2016). Whilst researchers have put forward the argument that DI and PT can complement each other to create a superior instructional technology (Binder & Watkins, 1990; Desjardins & Solcum, 1993), further research is necessary to show the generalizability of an instructional fluency approach across different curriculum subjects to remediate children's skill deficits. Investigations in this area could validate the use of an instructional fluency approach in PRU settings to help children improve their academic performance.

Chapter 7：Discussion of thesis

Broad overview of thesis aims and empirical chapters

The overall aim of this thesis was to help contribute to the evidence base surrounding the use of PT and DI methods to improve children's mathematics/numeracy skills. Specifically, we explored how these methods can be used in educational provisions (including schools and PRUs) to help support children considered proportionally 'at risk' of poor academic outcomes. By evaluating the existing literature (Chapter 1) we were able to identify several research questions that had not been addressed within the current published literature. Chapter 2 outlined the aims, rationale, and research methods underlying each of the proceeding empirical chapters. This thesis includes three manuscripts that have either been submitted (Chapters 3 and 6) or prepared (Chapter 4) for publication, as well as a close-to-practice impact report (Chapter 5).

Chapter 3

Implementation science details that there can be a disparity between the outcomes of efficacy and effectiveness trials (Kraft, 2020; Thornicroft et al., 2011). This disparity is due to the challenges associated with implementing interventions when experimental control is reduced and/or a research team withdraws support. Data from Durlak and DuPre (2008) suggested that teachers find it difficult to implement evidence-based educational strategies under the real-world conditions of the classroom, resulting in smaller effect sizes for academic outcomes. The majority of existing SAFMEDS literature report the outcomes of efficacy designs; whereby a researcher with expertise in the methods delivered the intervention to ensure high levels of fidelity (see for example Casey et al., 2003; Chapman et al., 2005; Hartnedy et al., 2005). Beverley et al. (2016) published the only known study investigating teacher led implementation of the SAFMEDS strategy following training. Whilst the results of this research were positive, it was still unclear what role a researcher can play in providing teachers with implementation support. With a growing number of schools across North Wales adopting the SAFMEDS strategy to support children's fluency of mathematics skills (Tyler et al., 2019), we were keen to investigate the impact of low-intensity coaching during a teacher led SAFMEDS mathematics program (Chapter 3).

Following training for the SAFMEDS strategy, we randomly allocated teaching staff to one of the two trial arms (support or no support). Those in the support arm received three in-situ implementation support visits from the first author across the duration of the project, along with email contact if required. The first author tailored this support based on the individual needs of the teaching staff, their school, and the children they were working with. The results from this study suggested that providing teaching staff with implementation support has greater impact on children's fluency of arithmetic, when compared with the outcomes of children attending schools that received no implementation support (MFaCTs: Grades 1-2, $d = 0.23$; MFaCTs: Grades 3-5, $d = 0.25$). In line with Kraft's (2020) revised benchmarks for research interpreting outcomes from preschool to Key Stage 5 educational interventions, the effect of coaching on both MFaCTs measures would be considered large. In the context of other education research investigating coaching programs, our effect sizes are larger than average (see Kraft et al., 2018; Lynch et al., 2019). This coaching model may provide a promising option for researchers looking to provide implementation support for the SAFMEDS strategy at scale. It may also broaden options for teaching staff seeking further professional development opportunities.

Chapter 4

After investigating teacher led implementation further, we identified that there may be certain elements of a SAFMEDS program that teaching staff may have been unable to implement under the day-to-day conditions of the classroom. To further understand staff views and experiences of using the SAFMEDS strategy we disseminated an online qualitative survey (Chapter 4). We asked 55 members of teaching staff what they perceived the benefits of using the SAFMEDS strategy to be; what factors they think help and hinder the success of SAFMEDS sessions; if they saw any changes in children's behavior as a result of engaging in a SAFMEDS program; and the reasons why they would or would not use the strategy in future. We believed gaining insight into these aspects of social validity may contribute to the existing evidence base and provide researchers with some additional context when training and/or coaching teaching staff. Using thematic analysis, the first author was able to identify five themes across staff's responses: (1) factors that promote progress; (2) factors that limit progress; (3) impact of competition; (4) confidence; and (5) inherent advantages of the SAFMEDS strategy.

Within Chapter 4, we also reported the outcomes of interviews with 26 children who had used the SAFMEDS strategy to support their fluency of mathematics skills. Prior to this, no known research had conducted a robust evaluation of the social validity children associate with the SAFMEDS strategy. The first author asked each of the children whether they enjoyed using the strategy; if there were any elements of the strategy they did not enjoy; whether they thought it was useful; and whether they would like to use it outside of school. We believed insight into these areas may help researchers to promote positive elements of the strategy within teacher training. The children's responses may also provide avenues for future research investigating factors that affect engagement and progress during a SAFMEDS program. Analyses of these transcripts revealed five themes: (1) enjoyment; (2) data; (3) sense of achievement; (4) skills; and (5) home use.

Chapter 5

The SAFMEDS strategy also gained traction within Inner City London, with XLP (a youth charity) promoting its use to help children improve their fluency of basic literacy and numeracy skills. Chapter 5 outlined a close-to-practice impact report, detailing the collaboration between researchers at Bangor University, staff at XLP, and educational provisions throughout the XL-LAN project. Through the use of exploratory analysis and descriptive statistics we were able to gauge the average fluency progress that children make across their engagement with the XL-LAN project, as well as the engagement at a provision-level. For numeracy SAFMEDS packs, children answered an average of 1.97 more cards correctly each session. When starting their first pack of numeracy cards, children could typically answer 26% of the cards in the pack correctly (~15.60 cards). This increased to 52% (~31.20 cards) on their personal best try. Children appear to make steady progress over the first 9 sessions for each pack, before their scores started to plateau (x 1 celeration). Moreover, the children in our sample spent a medium of 6.5 sessions on each numeracy pack, although this did range from 1 to 26 sessions.

In terms of the literacy SAFMEDS packs, children answered an average of 2.48 more cards correctly each session. In the beginning, children could typically answer 13% of the cards in the literacy packs correctly (~7.80 cards). This increased to 47% (~28.20 cards) on their personal best try. The children within our sample appeared to make consistent progress across sessions when practicing literacy skills; completing a median of 6 sessions on each literacy pack (range: 1 to 35 sessions). Variability across literacy and/or numeracy scores may extend to the delivery of the program across different provisions/settings. However, we need further data on implementation fidelity and demographic characteristics to enable us to complete a more in-depth evaluation.

Following the pilot of the XL-LAN project, we interviewed 38 children who had used the SAFMEDS strategy. We hoped that these interviews would expand our knowledge of how useful children perceive the strategy to be, whether they enjoy using it, and whether they would electively use it outside of a school setting. This extended the research that we conducted within Chapter 4, with particular emphasis on older children; including a sample of children preparing to sit their GSCE examinations. By applying thematic analysis to the interview transcripts, we identified five themes: (1) procedure, (2) improvement, (3) revision tool, (4) home use, and (5) withdrawal from class.

Chapter 6

DI programs can help children to acquire new skills and knowledge, whilst PT offers the tools for fluency assessment and data-driven decision making (Engelmann et al., 1988; Kubina & Yurich, 2012). Binder and Watkins (1990) argued that combing PT and DI together could create a superior instructional technology, which could help children make unprecedented academic gains. The existing literature exploring the use of an instructional fluency approach largely focuses on literacy skills, with no known published studies reporting its application to mathematics.

In Chapter 6, we used the CM program over 6 instructional weeks to help teach addition skills to five children who attended a PRU. The first author supplemented each CM lesson with fluency practice sheets and made data-driven decisions using a SCC (in line with PT methods). Each of the children attended a different number of sessions to due to reasons such as conflicting school events, illness, and refusal to comply (i.e., lack of assent). Contextually, this information is important when considering the outcomes of the follow-up measures. For example, Will completed 16.5 instructional fluency sessions, finishing 16 lessons from the CM addition program. His age-equivalence score on the TEMA-3 increased by 1-year 3-months and his standard mathematics score on the WRAT-4 improved by 7. In contrast, Leo completed 3.5 instructional fluency lessons, finishing 3 lessons from the CM addition program. His age-equivalence scores on the TEMA-3 did not change from baseline to follow-up and his standard mathematics score on the WRAT-4 reduced by 3. The SCCs for each child illustrate their fluency progress on the addition practice sheets and when the first author altered the skill level based on their data. The quantitative results from this small-scale pilot study suggest that the instructional fluency approach can help children attending PRUs to improve early mathematics skills, but this is contingent on their attendance to the sessions and engagement with the content.

Implications and applications

Throughout each chapter we have discussed several implications and applications of addressing each of our research questions. Below we aim to summarize the context and implications of this research for children, teachers, and researchers.

SAFMEDS training

The c-RCT that we presented within Chapter 3 helped us to understand how we can better support children following teacher training for a SAFMEDS program. By providing low-intensity coaching (three 1-hour in-situ support visits and email contact), the researcher was able to support teaching staff to implement the program. As a result, children attending schools where their teacher/TA received coaching made more fluency progress than their peers attending schools who did not receive coaching. This may have implications with regards to the professional development model that we offer schools as part of a SAFMEDS training package. The EEF (2019) have highlighted that training is a way of helping teachers to develop their conceptual understanding of an intervention but alone may not lead to effective implementation in the classroom. In their recent implementation guidance report, the EEF identified the importance of reinforcing initial training for interventions with expert follow-on support within school. The findings from our c-RCT suggest that this guideline may extend to the use of the SAFMEDS strategy, particularly within the context of supporting children who are performing below age-expected norms on mathematics measures.

The results from Chapters 4 and 5 suggest that there are certain challenges associated with implementing the SAFMEDS strategy under the day-to-day conditions of the classroom. For example, teaching staff indicated that cheating can occur within group based SAFMEDS sessions. This has implications in practice because it may result in children plotting invalid data on their SCCs and affecting subsequent data-driven decisions. To overcome this, trainers/coaches may need to integrate some of the following guidelines into the initial training for the SAFMEDS strategy and/or follow-on support packages. First, we should consider how group dynamics may foster unhealthy peer competition, and how this may contribute to individual children feeling embarrassed or lacking confidence if their scores are low. It may be appropriate to encourage children to keep their scores to themselves from the beginning of the program. Second, we should consider setting children individualized score targets on a regular basis. To do this, we could use trajectory lines on the SCC to help set realistic and attainable daily targets for each child (Johnson & Street, 2012) or encourage children to try and beat their personal best score each session (Hughes et al., 2007). Wyse (2001) provided some additional advice surrounding effective target setting, including encouraging children to reflect on their progress regularly. The SCC is a cardinal feature of the PT approach and is a useful tool to help children see where they started and how far they have come (Lindsley, 1995). Turning children's attention towards their own progress may help to limit their engagement with unhealthy peer competition and cheating tactics (Hughes et al., 2007). Third, if cheating is a persistent problem, we may want to consider scaling down the program to work with smaller groups of children. This will help to increase the one-on-one guidance we can offer to children during the error correction stage to support their skill development and increase the time we can allocate to validating their scores (e.g., we can work one-on-one with each child for at least one timing per session). If scaling down is not possible due to the number of children the school want/need to support using the SAFMEDS strategy, then we should try to validate each child's score on a semi-regular basis (e.g., watch each child complete a timing once a week) so that we have an accurate 'anchor' point to compare future scores with. Importantly, we should ask children to complete an observed timing when their SCC suggests that they have reached their fluency aim for a given skill. This will help to ensure that children do not move onto more difficult packs without mastering the necessary prerequisite skills.

During the early phases of implementation, teaching staff found that it took time for the children in their group to get used to using the cards and following the prescribed stages of the SAFMEDS strategy (Chapter 4). It might be useful for researchers/coaches to follow-up with schools during this early transitionary period after training to ensure that both staff and children are following the stages as closely to the intended design of the program as possible (i.e., have high levels of implementation fidelity). This may also provide a good opportunity to discuss how easily the school have found it to integrate the SAFMEDS sessions into their timetable. Without a sufficient number of opportunities each week to practice and assess their skills, children may make limited progress and schools may not see the results from the program that they were hoping for (Graf & Auman, 2005). Durlak and DuPre (2008) highlighted that low-quality implementation of educational interventions can lead to smaller effect sizes for outcome variables. As such, by engaging in discussions about implementation challenges early on, teaching staff and researchers can work together to find ways of more readily integrating the necessary stages of the SAFMEDS program more accurately into their routine.

Social validity

To our knowledge, Chapters 4 and 5 present the first qualitative explorations of key stakeholder views of the SAFMEDS strategy. Wolf (1978; as cited in Schwartz & Baer, 1991) wanted to make researchers aware that non-acceptance of interventions by key stakeholders could lead them to reject a program. Thornicroft et al's (2011) conceptualization of different types of evidence identified that stakeholders do not always continue using an intervention after the termination of a research study. As such, our research feeds directly into the implementation/maintenance phase of evidence, whereby we aimed to identify factors that might facilitate or prevent the uptake of the SAFMEDS strategy by teachers/TAs and children.

Approximately 90% of the children in Hunter et al's (2016) sample (*n* = 19) indicated that they would like to continue using the SAFMEDS strategy in school to improve their mathematics skills. However, Hunter et al. did not explore the reasons underlying *why* some children would continue to engage with it, whilst others would not. The majority of the children in our sample (Chapter 4 and 5) claimed that they enjoyed using the SAFMEDS strategy. They found it fun and appreciated the independence that it afforded them. Repeatedly, children reported that they were motivated by their scores improving and that they had noticed their ability to answer increasingly more difficult questions as the sessions and packs progressed. The majority of children that we interviewed also indicated that they would like to use the cards at home if their school provided the materials. This would provide them with the opportunity to continue improving their skills and allow them the opportunity to engage their family in the process (e.g., getting their parents to help or completing simultaneous timings with their siblings).

The teachers and TAs who completed the online survey (Chapter 4) noticed that the children in their group had gained confidence across the duration of the SAFMEDS program. This confidence extended into the classroom, where the children began raising their hands to answer more questions in class. Despite facing some additional challenges integrating the program into their school's timetable and overcoming cheating, the majority of teaching staff who completed our survey indicated that they would like to continue using it in future. This was largely due to the inherent advantages of the SAFMEDS strategy, such as the 1-minute sprints and the children becoming more fluent at basic skills.

Enhancing a research culture

Close-to-practice research helps to develop a research culture within organizations by supporting teams to think critically about gaps in their knowledge and services, use appropriate methods to elicit change, and evaluate outcomes (Cooke, 2005). BERA identified that educational research should focus on addressing problems in practice and encouraging researchers to develop working partnerships with key stakeholders within education to support PD (Wyse et al., 2018). PT methods lend themselves well to close-to-practice research due to the convention of using the SCC to make decisions about future instruction. When defining the PT approach, Lindsley (1990) was keen to put science into the hands of the learners and teachers who use it.

Chapter 3 demonstrates one example of how we worked with an educational organization (GwE) and schools to enhance the engagement and uptake of the SAFMEDS strategy across North Wales. GwE and the research team organized the recruitment of participating schools and the initial training session. During this training session, teachers and TAs could develop their pedagogical understanding of PT, the SCC, and the SAFMEDS strategy. They could then take this knowledge forward to support the children that they worked with. For schools allocated to the support arm of the trial, the first author aimed to provide some additional coaching that would in turn help teaching staff to implement the program to higher levels of fidelity. Each support visit provided a platform for teachers to reflect on their practice, gain advice from someone with prior experience and knowledge of the strategy, and discuss children's data. Following the results of this trial, we created a jargon-free impact report in collaboration with GwE explaining the outcomes of the trial and consequent recommendations for schools. Engagement with this training and strategy implementation can also help schools make effective use of their Pupil Deprivation Grant (PDG) from the Welsh Government (2013b). The PDG aims to help children from disadvantaged backgrounds overcome some of the additional barriers that prevent them from achieving their full protentional (including children who are eFSM). By informing schools of effective low-cost implementation strategies, educational organizations, and universities can help schools to invest in impactful interventions that support a greater number of students each year. As Kraft (2020) noted, stakeholders often view positive effect sizes from lower-cost interventions (such as the SAFMEDS strategy) as more favorably than more costly alternatives. We acknowledge that a full economic costing associated with using the SAFMEDS strategy in schools is necessary to validate this claim and compare it against the effects of other fluency-based mathematics interventions.

Chapter 5 provides an additional example of a close-to-practice project. Here, researchers at Bangor University, a youth charity (XLP), and educational provisions within inner city London collaborated to support children to develop basic literacy and numeracy skills. This project adopted a ‘train the trainer’ approach to increase the capacity for SAFMEDS implementation and support across inner city London. Supervisors and the XL-LAN coordinators became familiar with the process of reviewing data regularly and were able to identify invalid scores. As researchers, we were able to analyze the data in more depth to evaluate the program outcomes. This led to the generation of an impact report for funders and a simplified report for public dissemination.

Supporting children in PRUs

Previous statistics suggest that children who attend PRUs are ‘at risk’ of underperforming in mathematics (Ofsted, 2016), with recent observations from Estyn (2017) suggesting that some children attending these units do not develop the numeracy skills they need to support their learning across the curriculum. In an effort to engage children attending PRUs in scaffolded and unambiguous mathematics sessions to teach them basic arithmetic skills, we piloted the use on an instructional fluency program (using a combination of DI and

PT). Chapter 6 details the outcomes of this pilot, which focused on a sample of five boys who attended the unit on either a full time or dual time basis. In the short term, this program appeared to help four of the boys (Tom, Dean, Will, and Chris) to improve their early mathematics skills (assessed using the TEMA-3) beyond what would be expected over six school weeks. These children completed between 9 and 16 CM addition lessons, with accompanying fluency practice using randomized worksheets. Their age equivalence scores improved by 3 to 15 months over the six-week intervention period. Due to the small sample size included within this study further replications are necessary to establish the reliability of the results (Graham et al., 2012). Whilst further evaluation is needed, our findings suggest that the CM program could be a promising intervention approach to improve the mathematics attainment of children in PRU.

Dissemination

Edwards (2015) explained that researchers can disseminate their research through a range of approaches; with the most popular being through publications in journals and presentations at professional meetings or conferences. By engaging with dissemination efforts, we can help inform key stakeholders (i.e., those who will contact and use the intervention) and other researchers in the field of the outcomes of our trials. Longer term, this may affect professional practice and help to evolve the evidence base for educational interventions. Over the course of this PhD, we have worked towards disseminating the outcomes of each study with a range of audiences including professionals with a specific interest in ABA; professionals with an interest in general education pedagogy; and teaching staff. Table 7.1 outlines some of these dissemination efforts. We have also made an effort to provide simplified impact reports to the schools who contributed to each research study and we have integrated our findings into training for the SAFMEDS strategy.

Table 7.1

Dissemination Activities

Chapter reference	Conference	Publication status
3	• Experimental Analysis of Behaviour Group, London (2019). Oral presentation. • British Education Research Association, Swansea (2018). Poster presentation. • European Association for Behaviour Analysis, Würzburg (2018). Oral presentation. • A Child's World – New Shoes New Direction, Aberystwyth (2018). Oral presentation.	Under review
4	N/A	In preparation
5	• South West Doctoral Training Partnership Beyond Research: Society, Collaboration, and Impact (2018). Keynote presentation.	N/A
6	N/A	Accepted/In press

Limitations and future research

Chapter 3

One of the main limitations of the c-RCT (Chapter 3) was the lack of data relating to how coaching affected teachers/TAs fidelity of implementation across the program. Whilst we appreciate that this would have enabled us to draw stronger conclusions to support our theory of action, we were unable to collect these data due to practical and funding constraints. The first author conducted the support visits and therefore knew which trial arm each school had been allocated to. In order to reduce bias, we would have had to source an independent researcher to visit each of the 64 schools included in the randomization. This was logistically challenging given the geographical locations of each school, the times/days each school ran their SAFMEDS session, and the need to observe a session after each of round of support visits. Going forward, it may be more viable to assess changes in implementation fidelity over time across trial arms in a smaller-scale c-RCT.

The coaching element of this trial may have acted on two mechanisms: fidelity and/or accountability. In retrospect, we should have provided an anonymous survey to teachers in the support arm during their final visit to gain further insight into the role that personal accountability might have played. For example, we could have used Likert scales and asked teaching staff to rate statements such as "I would have felt embarrassed if the researcher had noticed I had not made a change to the SAFMEDS sessions that she had previously suggested" and "I felt proud receiving complements from a researcher on how I was running the SAFMEDS sessions" from "strongly disagree" to "strongly agree". Alternatively, we could have disseminated a qualitative survey to gain further insight into the thoughts and feelings that they experienced as a result of the researcher coming in for the support visits.

Chapter 4

Qualitative research does not happen in a vacuum. The research that we design and the questions that we ask are, in part, driven by both who we are as a person (*personal reflexivity*) and what we know from our own experience of theory and practice (*epistemological reflexivity*). As such, researchers should reflect upon the ways in which they may have unintentionally implicated the research and its findings (Willig, 2008). However, it should be noted that reflexivity does not discredit the usefulness of the findings of qualitative research. Rather, this exercise serves to add a layer of transparency to the research process. Tuval-Mashiach (2017) suggested that researchers consider three reflexive statements: 'what we did', 'how we did it', and 'why we did it'.

We discussed our research aims, as well as the context and the justification for using qualitative research and thematic analysis in detail in Chapter 2. This addresses the 'what we did' and 'how we did it' portions of Tuval-Mashiach's (2017) model of transparency. With regards to ‘why we did it’, the first author had several years’ experience of using the SAFMEDS strategy in schools. Since 2016, she had been working closely with teachers to support implementation (including the c-RCT presented within Chapter 3) and had noticed some recurring challenges that they faced after initial training. This included children cheating and experiencing difficulty using the SCC. Several schools that she had worked with in the past had opted not to use the SCC and thus were making decisions based on children’s tabulated data, if they reviewed it at all. This sometimes meant that children would go prolonged periods of time making little to no progress and thus compromising the fidelity of the PT/SAFMEDS approach. The first author was interested in turning some of these anecdotal observations into a more robust evaluation of teacher’s/TAs views and experiences of using the SAFMEDS strategy in schools. When developing the questions, the first author was keen not to prime specific responses (e.g., relating cheating and use of the SCC) in case this deflected from some of the other challenges that teachers faced when implementing the program. Additionally, the first author was keen to capture whether teachers felt like there were associated benefits with using the SAFMEDS strategy with the children that they supported.

The staff survey did indicate that cheating was a common problem that they faced during a SAFMEDS program and tended to stem from unhealthy peer competition in group settings. The responses to this survey also highlighted several other challenges that teachers faced throughout a SAFMEDS program, including the initial transitionary period and integrating the sessions into the school’s timetable. The use of SCCs was not something that came up in the survey responses. This might have been because we asked the wrong questions, or it is not something that teachers consider to be a salient challenge for them. This may be worth investigating more directly in a further qualitative study.

In her past experience, the first author had noticed that some children enjoyed engaging with the SAFMEDS strategy (e.g., they would ask when it was their turn to come out of class, they would be excited to share their score with her if they had a new personal best). Conversely, she had come across some children who found it difficult to engage with the strategy (e.g., they would engage in with cheating tactics or behaviors that challenge during the SAFMEDS session). During her support visits throughout the c-RCT (Chapter 3), the teachers/TAs that she had been working with indicated similar experiences. This provided the initial motivation to conduct a more robust evaluation of children's experiences and views regarding the SAFMEDS strategy.

The first author was conscious of the age of the children in the sample, so started off with a yes/no statement to help facilitate the discussion (e.g., "do you enjoy using SAFMEDS?"). Once the children responded, she followed this up with an open-question prompt (e.g., "why do you enjoy it?"). Ultimately, she wanted to gain insight into some of the underlying factors that might affect children's engagement with the strategy.

This study provided some useful insight into teacher's and children's experiences of using the SAFMEDS strategy in schools. It prompted a useful discussion with regards to the challenges that teachers may face whilst implementing the strategy as well factors that might promote/hinder children's engagement and performance during sessions. That said, the research methods that we used may have influenced that themes that we were able to identify. We used an online survey to collect teacher's views surrounding implementation and perceived benefits. This enabled us to recruit a larger sample than we possibly could have by running interviews or focus groups (Van Selm & Jankowski, 2006) but may have led us to collect data that was not as rich. Using surveys meant that the first author was unable to ask for elaboration if a teacher had provided an interesting response that could have led to an insightful discussion. By having these discussions, we may have generated a richer data set and further themes.

Kortesluoma et al., (2003) explained that using interviews with children can be a favorable method of collecting qualitative data over surveys because it allows them to have space and time to discuss their views and experiences, without limiting their responses to a narrow range of categories. Moreover, interview responses are also not bound by children's written comprehension abilities. We opted to interview the children to overcome these two issues and generate what we thought would be a richer and more insightful data set. However, it is possible that running the interview on a one-to-one with the researcher (who they knew was there to talk to them about the SAFMEDS strategy) prompted some of the children to provide responses that they deemed to be socially desirable. All of the children that we interviewed for this study ($N = 26$) indicated that they enjoyed using the SAFMEDS strategy, however we acknowledge that this finding might be an over-representation. We may also have been able to identify further themes relating to elements of the strategy that children find difficult or unenjoyable if we had employed a research method that afforded a level of anonymity (e.g., a simple survey).

In terms of future research, the interviews with children in Chapter 3 indicated that it might be worth exploring the use of SAFMEDS in a home-setting. The majority of the children in our sample indicated that they would be interested in using the cards at home if their school could provide them with the materials. To date, no known published research has reported the effects of using SAFMEDS as a homework strategy. This could be a useful way of increasing practice opportunities for children from demographic groups who may be disproportionally at risk of underachieving at school.

Chapter 5

Within Chapter 5 we highlighted some limitations of the XL-LAN pilot project that led to recommendations for future practice. One notable recommendation was the use of standardized achievement tests to help build a profile of the skills that each child has mastered and where they may need additional instruction and/or practice. Within session 1 of the program, children were able to answer an average of 26% of the cards correctly in their first numeracy pack and 13% of the cards correct in their first literacy pack. This may suggest either dysfluent performance or incorrect pack placement. Miller and Heward (1992) argued that children need to first learn *how* to perform a skill to a level of mastery before turning attention to fluency training. Using a standardized assessment such as the WRAT-4 (Wilkinson & Robertson, 2006) may aid more time-efficient placement onto more appropriate packs earlier in the program; rather than relying on slicing back skills several times in the early weeks.

The data from this study indicated that children completed an average of two SAFMEDS sessions per week; which is one day fewer than employed within previous research (e.g., Greene, Mc Tiernan, & Holloway, 2018; Hunter et al., 2016). SAFMEDS programs rely on regular practice to ensure that children are able to contact the material, address any “not yets”, and elicit retention (Graf & Auman, 2005); by engaging in the content sporadically and infrequently children may not make the desired progress. In future research, it will be useful to consider strategies to integrate SAFMEDS more readily into the school timetable particularly within a ‘train the trainer’ context. School-based provisions may need to consider alterations to the program such as: adopting a peer tutoring model (see for example, Greene et al., 2018); exploring the practicalities of the children using the SAFMEDS strategy at home during the week to increase the number of practice opportunities; and/or adjusting the dosage of SAFMEDS timings (e.g., completing one per day rather than four several times per week).

During the interviews, the children enrolled on the XL-LAN project highlighted the versatility and applications of the SAFMEDS strategy. This includes the benefits of using the strategy to support class tests and statutory GCSE exams. Future research should validate the use of the SAFMEDS strategy to support the revision of fact and definition-based GCSE content and the consequent effects on test performance. This would build upon the work of Hunter et al. (2018) and provide insight into how the SAFMEDS strategy may function given the increasing difficulty of GCSE content compared to the primary school curriculum the children practiced as part of the XL-LAN project.

Chapter 6

Due to the dual registration and absenteeism of some children it was difficult to estimate the number of hours they spent in the PRU compared to their mainstream school each week. At the beginning of the study, Tom, Will, and Leo were registered to attend the PRU on a dual registry basis so it also would have been beneficial to know if they were receiving any supplementary mathematics support in their mainstream school. Access to this information, if we were able to get it, may have assisted he interpretation of our results.

We opted to use a small *N* design for this project for several reasons. First, there was a limited sample of children available for this study. The PRU accommodated approximately 20 children, and only 10 returned parental consent forms. Second, the researcher conducted the sessions on a one-on-one basis, so she was unable to work with more than five children per week. Third, it was difficult to recruit a matched sample control group due to the children's ages, diagnostic labels, the percentage of time they attended the PRU, and their baseline mathematical abilities. And fourth, due to the time constraints surrounding this project we were also unable to explore the possibility of the children acting as their own controls and measuring their performance growth for a period of time before and after exposure to the intervention. Future extensions to this study should consider a feasible way to collect control data to establish whether the instructional fluency approach has any additional advantages for children attending PRUs, compared to the typical classroom instruction that they are exposed to. For example, researchers could employ a matched pairs design (with one group acting as a treatment as usual condition and the other receiving the intervention) or they could consider using children as their own controls (e.g., an AB design). Due to the lack of evidence investigating the use of DI and/or PT in a PRU, a final extension to this project would be to disentangle the two approaches. It would be useful to compare how DI and PT approaches work in isolation, as well as how the outcomes compare when they are combined into the instructional fluency approach.

COVID-19

In March 2020, schools across the UK temporarily closed due to the coronavirus pandemic (COVID-19). Whilst there is still limited data available to establish the impact that the school closures had on children's education, there is some evidence focusing on school closures over prolonged periods such as the summer holiday. The EEF (2020) conducted a rapid evidence assessment on the available literature to examine the potential impact that these closures might have had on the attainment gap within UK education. Their results indicated that prolonged school closures can widen the attainment gap between children from low socioeconomic backgrounds and their peers by between 11% and 75% (median = 36%). Their report also provided some guidance for schools to help remediate the effects of the unexpected school closures due to COVID-19. First, it has become increasingly important to ensure that remote learning opportunities consist of clear explanations, scaffolding, and feedback. Second, it is likely that the attainment gap will have widened when schools reopen in full capacity. Catch-up support will become increasingly important, including assessment of learning and targeted instruction.

Whilst it is unlikely that a single intervention will help remediate the impact of school closures, establishing a strong evidence base for educational interventions may help schools to make informed choices about which to use and how best to support implementation. Manualized DI programs are scripted to help aid scaffolded and explicit teaching (Engelmann, 2007a; Kozioff et al., 2000). The lessons also incorporate the model-lead-test-retest format to help correct errors and ensure regular review of concepts (Bechtolt et al., 2014). Likewise, PT offers a system for teachers and children to monitor progress and make data-informed decision about subsequent instruction (Kubina & Yurich, 2012; Lindsley, 1995). By applying this to the SAFMEDS strategy, children can practice and assess fact-based skills to fluency (Graf & Auman, 2005; Johnson & Layng, 1992). The research that we have presented within this thesis provides insight into the effects of PT and DI methods with children from ‘at risk’ groups, including children who are underachieving in numeracy and children who attend a PRU.

The data that we have presented suggests that children can use the SAFMEDS strategy to improve their fluency of basic mathematics and literacy skills (Chapters 3-5). Moreover, children can complete the procedure independently (Chapter 4) and some would like to use the strategy at home if they had the necessary materials (Chapters 4 and 5). By increasing the number of practice opportunities at home, schools may be able to help children catch-up and reduce the attainment gap.

DI manuals come with scripted lessons which aim to teach children new skills in the most time-efficient way possible (Engelmann et al., 1988). McGraw-Hill (2019) market the CM program as a remedial intervention to help children catch up with their age-matched peers. Going forward, teachers and researchers may want to explore the benefits of using DI programs both in schools and/or PRUs and remotely to help children to master rules, concepts, and strategies across different core subjects/topics.

Conclusion

The four empirical chapters within this thesis aimed to develop the evidence base surrounding the implementation of PT and DI approaches to support children's development of basic mathematics skills. In Chapters 3, 5, and 6 we placed focus on supporting the academic outcomes of children who are considered disproportionally more likely to underachieve at school. This included: children eligible for free school meals; children unable to perform age-expected mathematics skills; children from areas of high social deprivation within inner city London; and children who attend a PRU.

Despite some challenges that may arise with implementation there are several clear benefits associated with children using the SAFMEDS strategy in schools, including increased fluency of basic skills, improved confidence in their mathematical abilities, and having a visual record of their progress. We established that teachers and TAs benefit from implementation support after initial training for the SAFMEDS strategy. By engaging in low-intensity coaching, teachers can support the children in their group to make greater fluency progress. We have also shown, through the use of qualitative methods, that the approach is generally well received by teachers.

In an effort to understand how we might raise numeracy standards in PRUs, we conducted a small-scale pilot study of an instructional fluency program. We used the CM addition curriculum and individualized fluency practice sheets over six-instructional weeks on a one-on-one basis. The results suggested that if we can encourage children to engage with the strategy, it is possible for them to improve early-mathematics skills.

Reflection and future aims

As I (first author) have progressed through this PhD, I have had the pleasure of interacting with lots of teachers, children, and colleagues— each teaching me so much about what it really means to be a researcher. It has been a pleasure being out in the field to deliver training and support to those who have gone forward to use these methods. Each school visit made me reflect on just how powerful these strategies can be if they are used correctly. I am so grateful to have been able to see first-hand the positive effect that these strategies can have. Knowing that I have contributed in some way to helping these children achieve goals, and gain confidence in mathematics, will always make this job worth doing.

The data that makes up the empirical chapters of this thesis has gone some way towards enhancing evidence-based practice in schools. This in turn contributes to some of the goals set out by wider educational initiatives, such as those proposed by the Welsh Government, the EEF, and BERA. The findings from this thesis have enabled us to adapt the training and support that we are able to offer to schools in the region, and we hope that our dissemination efforts have gone some way towards improving the uptake and quality of implementation on a wider scale.

In July 2019, I was fortunate enough to receive a scholarship to attend the Summer School Institute at Morningside Academy in Seattle. This trip was instrumental in supporting my development and knowledge of how evidence-based practices can support skill development across both numeracy and literacy curricula. The classroom practices instilled at Morningside Academy follow a model of generative instruction, which builds on research surrounding content analysis, instructional design, program placement, classroom organization and management, critical thinking, reasoning, and self-regulated decision-making (Johnson & Street, 2004). Being mentored by some of the leading pioneers in this field provided me with the knowledge and confidence I needed to take to the front of a class and deliver lessons for myself. I now have a greater knowledge of how PT, DI, and think aloud problem-solving strategies can be used to support children's development of key academic skills.

The end of my PhD coincides with the COVID-19 pandemic. Due to the prolonged period of school closures, there is a concern that a growing number of children will need support to catch-up with age-expected norms (EEF, 2020). Despite children returning to schools, the current COVID-19 message in Wales prevents us from delivering face-to-face training and support visits. Over the coming months, I will be working with GwE (the Regional School Effectiveness and Improvement Service for North Wales) to help schools adopt methods that will accelerate children's learning. This includes creating new SAFMEDS packs to support the national numeracy curriculum and developing resources to promote the use of the SAFMEDS strategy at home. We will also be exploring ways to deliver training and support schools remotely. Longer term, I would also like to build upon the work detailed in Chapter 6 and help develop understanding of how DI and PT can be used to support children who attend PRUs. This extension will include gaining control data and upskilling TAs to deliver the program.

Lindsley (1990) was keen to put science into the hands of the teachers. This idea is growing in increasing relevance and importance; both in the context of the effects of COVID-19 on the education system but also the up-scaling of intervention uptake. The journey that led me to write this thesis has taught me the benefits of sharing knowledge about the use of PT and DI, as well supporting teachers to be able to use these methods with the children that they work with. Every child has the ability and right to learn, and we have the ability to provide teachers with the right methods and philosophies to achieve just that.

We will take you where you are, and we'll teach you. And the extent to which we fail is our failure, not yours. We will not cop out by saying "he can't learn". Rather, we will say, "I failed to teach him. So, I better take a good look at what I did and try and figure out a better way". (Engelmann, 2007b)

References

Adams, G. L., & Engelmann, S. (1996). *Research on DI: 25 years beyond DISTAR.* Seattle: Educational Achievement Systems.

Adda Ragnarsdóttir, G. (2007). Teaching an Icelandic student with autism to read by combining direct instruction and precision teaching. *European Journal of Behavior Analysis, 8*(2), 289-304. doi: 10.1080/15021149.2007.11434290

Aninao, T., Acevedo, D., Newsome, K., & Newsome, D. (2015). Putting the precision in precision teaching: using the standard celeration chart for ongoing assessment. *Behavioral Development Bulletin, 20*(2), 168-172. doi: 10.1037/h0101312

Archer, A., & Hughes, C. (2011). *Explicit Instruction: Efficient and effective teaching.* New York: Guilford Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. doi: 10.1016/j.jml.2007.12.005

Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 20*, 313-327.

Banerjee, P. A. (2016). A systematic review of factors linked to poor academic performance of disadvantaged students in science and maths in schools. *Cogent Education, 3*(1), 1-17. doi: 10.1080/2331186X.2016.1178441

Barnett, S. D., Heinemann, A. W., Libin, A., Houts, A. C., Gassaway, J., Sen-Gupta, S., ... & Brossart, D. F. (2012). Small N designs for rehabilitation research. *Journal of Rehabilitation Research & Development*, *49*(1), 175-186. doi: 1682/JRRD.2010.12.0242

Barriball, K. L., & While, A. (1994). Collecting data using a semi-structured interview: a discussion paper. *Journal of Advanced Nursing-Institutional Subscription*, *19*(2), 328-335.

Bechtolt, S., McLaughlin, T. F., Derby, K. M., & Blecher, J. (2014). The effects of direct instruction flashcards and a model, lead, test procedure on letter recognition for three preschool students with developmental disabilities. *Journal on Developmental Disabilities*, *20*(1), 5-15.

Beverley, M., Hughes, J. C., & Hastings, R. P. (2009). What's the probability of that? Using SAFMEDS to increase undergraduate success with statistical concepts. *European Journal of Behavior Analysis, 10*(2), 235-247. doi: 10.1080/15021149.2009.11434321

Beverley, M., Hughes, J. C., & Hastings, R. P. (2016). Using SAFMEDS to assist language learners to acquire second-language vocabulary. *European Journal of Behavior Analysis, 17*(2), 131-141. doi: 10.1080/15021149.2016.1247577

Beverley, M., Hughes, J. C., & Hastings, R. P. (2018). Improving essential numeracy skills in primary school children using a brief fluency-building intervention: A randomised control trial. *Wales Journal of Education, 20*(1), 114-134. doi: 10.16922/wje.20.1.7

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst*, *19*(2), 163-197. doi: 10.1007/BF03393163

Binder, C. (2003). Doesn't everybody need fluency? *Performance Improvement, 42*(3), 14.20.

Binder, C. (2010). Building fluency performance: Measuring response rate and multiplying response opportunities. *The Behavior Analyst Today, 11*(4), 214-225

Binder, C., Haughton, E., & Bateman, B. (2002). *Fluency: Achieving true mastery in the learning process.* Retrieved from https://www.researchgate.net/profile/Ogholgol_Nazari/publication/269793677_On_Situating_the_Stance_of_Practice_in_SLA/links/56b44f9808ae61c480581e9d.pdf [Accessed 7 September 2020]

Binder, C., & Watkins, C. L. (1990). Precision teaching and direct instruction: Measurably superior instructional technology in schools. *Performance Improvement Quarterly, 3*(4), 74-96.

Blackwell, A., Stookey, S., & McLaughlin, T. F. (1996). The effects of using direct instruction and a re-reading contingency with precision teaching. *Journal of Precision Teaching and Celeration*, *13*(2), 19-22.

Bliss, S. (2006). Test of Early Mathematics Ability—Third Edition. *Journal of Psychoeducational Assessment, 24*(1), 85-91. doi: 10.1177/0734282905282839

Bolich, B., & Sweeney, W. J. (1996). An eleven-year old girl's use of repeated reading SAFMEDS and see/write-think/write practice to develop fluent reading in Hebrew. *Journal of Precision Teaching and Celeration, 14*(1), 57-71.

Borm, G. F., Melis, R. J. F., Teerenstra, S., & Peer, P. G. (2005). Pseudo cluster randomization: a treatment allocation method to minimize contamination and selection bias. *Statistics in Medicine, 24*(23), 3535-3547. doi: 10.1002/sim.2200

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*(2), 125-230. doi: 10.3102/00346543073002125

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage.

Boyce, T. E. (2003). Moving from precision teaching to precision measurement: We need to just DO IT. *European Journal of Behavior Analysis, 4*(1-2), 59-63. doi: 10. 1080/15021149.2003.11434216

Brady, K. K., & Kubina, R. M. (2010). Endurance of multiplication fluency for students with attention deficit hyperactivity disorder. *Behavior Modification, 34,* 79-93. doi: 10.1177/0145445510361331

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi: 10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2012). *Thematic analysis.* In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbooks in psychology®. APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (p. 57–71). American Psychological Association.

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Croydon, United Kingdom: Sage.

Calkin, A. B. (2002). Estimated number of standard celeration charts as of 2000. *Journal of Precision Teaching & Celeration, 18*(1), 33-34.

Calkin, A. B. (2005). Precision Teaching: The Standard Celeration Charts. *The Behavior Analyst Today, 6*(4), 207-215.

Carnine, D. W. (1976). The effects of two teacher presentation rates on off task behavior, answering correctly, and participation. *Journal of Applied Behavior Analysis, 9*(2), 199-206. doi: 10.1901/jaba.1976.9-199

Casey, J., McLaughlin, T. F., Weber, K. P., & Everson, M. (2003). The effects of five-minute practice, unlimited practice, with SAFMED cards on correct and error rate in math facts for two elementary school children with learning disabilities. *International Journal of Special Education*, *18*(1), 66-72.

Cates, G. L., & Rhymer, K. N. (2003). Examining the relationship between mathematics anxiety and mathematics performance: An instructional hierarchy perspective. *Journal of Behavioral Education, 12*(1), 23-34. doi: 10.1023/A:1022318321416

Celik, S., & Vuran, S. (2014). Comparison of direct instruction and simultaneous prompting procedure on teaching concepts to individuals with intellectual disability. *Education and Training in Autism and Developmental Disabilities, 49*(1), 127-144.

Centre for the Use of Research & Evidence in Education (2011). *Report of Professional Practitioner Use of Research Review: Practitioner Engagement in and/or with Research*. Retrieved from http://www.curee.co.uk/files/publication/[site-timestamp]/Practitioner%20Use%20of%20Research%20Review%20-%20FINAL%2011_02_11.pdf [Accessed 22 July 2020]

Chapman, S. S., Ewing, C. B., & Mozzoni, M. P. (2005). Precision teaching and fluency training across cognitive, physical, and academic tasks in children with traumatic brain injury: A multiple baseline study. *Behavioral Interventions, 20*, 37-49. doi: 10.1002/bin.168

Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283-292. doi: 10.3102/0013189X16656615

Chiesa, M., & Robertson, A. (2000). Precision teaching and fluency training: Making maths easier for pupils and teachers. *Educational Psychology in Practice, 16*(3), 297-310. doi 10.1080/713666088

Cihon, T. M., Sturtz, A. M., & Eshleman, J. (2012). The effects of instructor-provided or student-created flashcards with weekly, one-minute timings on unit quiz scores in introduction to applied behavior analysis courses. *European Journal of Behavior Analysis*, *13*, 47-57. doi: 10.1080/10521149.2012.11434404

Clements, D. H., Sarama, J., Wolfe, C. B., & Spilter, M. E. (2015). Sustainability of a scale-up intervention in early mathematics. A longitudinal evaluation of implementation fidelity. *Early Education and Development, 26*, 427-449. doi: 10.1080/10409289.2015.968242

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Ed.). New York: Erlbaum.

Connolly, P., Biggart, A., Miller, S., O'Hare, L., & Thurston, A. (2017). *Using randomised controlled trials in education*. United Kingdom: Sage.

Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children, 72*(2), 135-144. doi: 10.1177/001440291307900201

Cooke, J. (2005). A framework to evaluate research capacity building in health care. *BMC Family Practice*, *6*(44). doi: 10.1186/1471-2296-6-44

Cunningham, D., McLaughlin, T. F., & Weber, K. P. (2012). The use and evaluation of verbal prompting with see to say problems and answers and SAFMEDS to teach math facts to a student with learning disabilities. *International Journal of Advances in Psychology, 1*(2), 37-39.

Dalton, D. W., & Hannafin, M. J. (1988). The effects of computer-assisted and traditional mastery methods on computation accuracy and attitudes. *The Journal of Educational Research, 82*(1), 27-33. doi: 10.1080/00220671.1988.10885861

Datchuk, S. M. (2017). A direct instruction and precision teaching intervention to improve the sentence construction of middle school students with writing difficulties. *The Journal of Special Education*, *51*(2), 62-71. doi: 10.1177/0022466916665588

Debski, R., & Gruba, P. (1999). A qualitative survey of tertiary instructor attitudes towards project-based CALL. *Computer Assisted Language Learning*, *12*(3), 219-239. doi: 10.1076/call.12.3.219.5715

Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behaviour. *Journal of Personality and Social Psychology, 53*(6), 1024-1037. doi: 10.1037/0022-3514.53.6.1024

Dell, C. A., Harrold, B., & Dell, T. (2008). Test review: Wilkinson, G.S., & Robertson, G. J. (2006). Wide Range Achievement Test—Fourth Edition. *Rehabilitation Counselling Bulletin, 52*(1), 57-60. doi: 10.1177/0034355208320076

Delli Sante,A., McLaughlin, T. F., & Weber, K. P. (2001). The use and evaluation of a direct instruction flash card strategy on multiplication math facts mastery with two students with developmental disabilities and attention deficit hyperactivity disorder. *Journal of Precision teaching and Celeration, 6*, 45-46.

Demie, F. & Strand, S. (2006). English language acquisition and educational attainment at the end of secondary school. *Educational Studies, 32*(2), 215-231. doi: 10.1080/03055690600631119

Department for Children, Schools and Families (2008). *Back on track: A strategy for modernising alterative provision for young people.*

Department for Education (2013). *Mathematics programmes of study: Key stages 1 and 2. National curriculum in England.* Retrieved online https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/335158/PRIMARY_national_curriculum_-_Mathematics_220714.pdf [Accessed 17 January 2020]

Department for Education (2016a). *Schools, pupils and their characters: January 2016.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/552342/SFR20_2016_Main_Text.pdf [Accessed 11 October 2018]

Department for Education (2016b). *Educational Excellence Everywhere*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/508447/Educational_Excellence_Everywhere.pdf [Accessed 25 August 2020]

Department for Education (2016c). *Achievement of 15-year-olds in England: PISA 2015 National Report.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/574925/PISA-2015_England_Report.pdf [Accessed 15 November 2019]

Department for Education (2017). *Alternative provision: Effective practice and post 16 transition.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/585550/Alternative_provision_effective_practice_and_post-16_transition.pdf [Accessed 3 August 2020]

Desjardins, E. A., & Slocum, T. A. (1993). Integrating precision teaching and direct instruction. *Journal of Precision Teaching, 10*(2), 20-24.

Desoete, A., Ceulemans, A., DeWeerdt, F., & Pieters, S. (2010). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study. *British Journal of Educational Psychology, 82*, 64-81. doi: 10.1348/2044-8279.002002

Donaldson, G. (2015). Successful futures: Independent Review of Curriculum and Assessment Arrangements in Wales. *Retrieved from* https://gov.wales/sites/default/files/publications/2018-03/successful-futures.pdf [Accessed 17 December 2019]

Dowker, A. (2016). Factors that influence improvement in numeracy, reading, and comprehension in the context of a numeracy intervention. *Frontiers in Psychology, 7*, 1929-1939. doi: 10.3389/fpsyg.2016.01929

Durlak, J. A. & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on programme outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3-4), 327-350. doi: 10.1007/s10464-008-9165-0

Education Endowment Foundation (2017). *Improving mathematics in key stages two and three guidance report.* Retrieved from https://educationendowmentfoundation.org.uk/public/files/Support/Publications/Maths/KS2_KS3_Maths_Guidance_2017.pdf [Accessed 17 December 2019]

Education Endowment Foundation (2019). *Putting evidence to work: A school's guide to implementation guidance report.* Retrieved from https://educationendowmentfoundation.org.uk/public/files/Publications/Implementation/EEF_Implementation_Guidance_Report_2019.pdf [Accessed 13 February 2020]

Education Endowment Foundation (2020). *Impact of school closures on the attainment gap: Rapid evidence assessment.* Retrieved from https://educationendowmentfoundation.org.uk/public/files/EEF_(2020)_-_Impact_of_School_Closures_on_the_Attainment_Gap.pdf [Accessed 10 Aug 2020]

Edwards, D. J., Dissemination of research results: on the path to practice change. *The Canadian Journal of Hospital Pharmacy, 68*(6), 465-469. doi: 10.4212/cjhp.v68i6.1503

Egbert, R. L. (1981). *Some thoughts about project follow through thirteen years later.* Retrieved from https://files.eric.ed.gov/fulltext/ED244733.pdf [Accessed 25 May 2020]

Engelmann, S. (1999). The benefits of direct instruction: Affirmative action for at-risk students. *Educational Leadership, 57*(1), 77-79.

Engelmann, S. (2007a). Student-Program alignment and teaching to mastery. *Journal of Direct Instruction, 7*(1), 45-66.

Engelmann, S. (2007b). *Teaching needy kids in our backwards system: 42 years of trying.* Eugene, OR: ADI Press.

Engelmann, S. (2014). *Successful and confident students with Direct Instruction*. Eugene, OR: National Institute For Direct Instruction (NIFDI) Press.

Engelmann, S., Becker, W. C., Carnine, D., & Gersten, R. (1988). The direct instruction follow through model: Design and outcomes. *Education and treatment of Children*, *11*(4), *303*-317.

Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications*. New York: Irvington Publishers.

Engelmann, S., & Carnine, D. (2005). *Corrective Mathematics* [intervention program].

Engelmann, S. E., Osborn, J., Haddox, P., & Hanner, S. (1978). *Corrective Reading Program, Comprehension A*. Chicago, IL: Science research associates [Instructional program]

Eshleman, J. W. (1985). Improvement pictures with low celerations: An early foray into the use of SAFMEDS. *Journal of Precision Teaching, 6*(3), 54-63.

Estyn (2015). *Education other than at school: a good practice survey.* Retrieved from: https://www.estyn.gov.wales/sites/www.estyn.gov.wales/files/documents/Education_other_than_at_school_a_good_practice_survey_-_June_2015.pdf [last accessed 27 November 2019]

Estyn (2017). *The Annual Report of Her Majesty's Chief Inspector of Education and Training in Wales.* Retrieved from https://www.estyn.gov.wales/sites/www.estyn.gov.wales/files/documents/Estyn_Annual%20Report_2018_English_Accessible.pdf [Accessed 21 June 2019]

Evans, M., & Boucher, A. R. (2015). Choice: Supporting student autonomy to foster motivation and engagement in learning. *Mind, Brain, and Education, 9*(2), 87-91. doi: 10.1111/mbe.12073

Fawcett, S. B. (1991). Social validity: A note on methodology. *Journal of Applied Behavior Analysis, 24*(2), 235-239. doi: 10.1901/jaba.1991.24-235

Feingold, A. (2015). Confidence interval estimation for standardized effect sizes in multilevel and latent growth modelling. *Journal of Consulting and Clinical Psychology, 83I*(1), 157-168. doi: 10.1037/a0037721

Fishel, M., & Ramirez, L. (2005). Evidence-based parent involvement interventions with school-aged children. *School Psychology Quarterly, 20*(4), 371-402.

Fletcher, S., & Mullen, C. A. (Eds.). (2012). *SAGE handbook of mentoring and coaching in education*. Los Angeles, California: Sage.

Flores, M. M., & Kaylor, M. (2007). The effects of a direct instruction programme on the fraction performance of middle school students at-risk for failure in mathematics. *Journal of Instructional Psychology, 34*(2), 84-94.

Foster, S. L. & Mash, E. J. (1999). Assessing social validity in clinical treatment issues and procedures. *Journal of Consulting and Clinical Psychology, 67*(3), 308-319.

Francis, B., Connolly, P., Archer, L., Hodgen, J., Mazenod, A., Pepper, D., ... & Travers, M. C. (2017). Attainment Grouping as self-fulfilling prophesy? A mixed methods exploration of self confidence and set level among Year 7 students. *International Journal of Educational Research*, *86*, 96-108. doi: 10.1016/j.ijer.2017.09.001

Fuller, J. L., & Fienup, D. M. (2018). A preliminary analysis of mastery criterion level: effects on response maintenance. *Behaviour Analysis in Practice, 11*(1), 1-8. doi: 10.1007/s40617-017-0201-0

Furlong, J. (2015). *Teaching tomorrow's teachers*. Retrieved from https://beta.gov.wales/sites/default/files/publications/2018-03/teaching-tomorrow's-teachers.pdf [Accessed 15 November 2019]

Gałecki, A., & Burzykowski, T. (2013). Linear mixed-effects model. In *Linear Mixed-Effects Models Using R* (pp. 245-273). Springer, New York, NY. doi: 10.1007/978-1-4614-3900-4

Gallagher, E. (2006). Improving a mathematical key skill using precision teaching. *Irish Educational Studies, 25*(3), 303-319. doi: 10.1080/03323310600913757

Gallagher, E., Bones, R., & Lombe, J. (2006). Precision teaching and education: Is fluency the missing link between success and failure? *Irish Educational Studies, 25*(1), 93-105. doi: 10.1080/03323310600597642

Geary, D. C. (2000). From infancy to adulthood: The development of numerical abilities. *European Child & Adolescent Psychiatry, 9*(2), 11-16. doi: 10.1007/s007887007004

Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement performance in mathematics. *Journal of Developmental and Behavioral Pediatrics, 32*(3), 250-263. doi: 10.1097/DBP.0b013e318209edef

Geiger, V., Goos, M., & Forgasz, H. (2015). A rich interpretation of numeracy for the 21st century: a survey of the state of the field. *ZDM Mathematics Education, 47*(4), 531-548. doi:10.1007/s11858-015-0708-1

Gelman, A. (2006). Multilevel (hierarchical) modeling: what it can and cannot do. *Technometrics*, *48*(3), 432-435. doi: 10.1198/004017005000000061

Ginsburg, H. P. & Baroody, A. J. (2003). *Test of Early Mathematics Ability, Third edition.* Austin, Texas: PRO-ED [measurement instrument].

Glang, A., Singer, G., Cooley, E., & Tish, N. (1991). Using direct instruction with brain injured students. *Direct Instruction News, 11*(1), 23-28.

Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health, 93*(8), 1261-1267. doi: 10.2105/AJPH.93.8.1261

Gorard, S. (2012). Who is eligible for free school meals? Characterising free school meals as a measure of disadvantage in England. *British Educational Research Journal, 38*(6), 1003-1017. doi: 10.1080/01411926.2011.608118

Gorard, S., Huat See, B., & Siddiqui, N. (2020). What is the evidence on the best way to get evidence into use in education? *Review of Education, 8*(2), 570-610. doi: 10.1002/rev3.3200

Graf, S. A., & Auman, J. (2005). SAFMEDS: A tool to build fluency. Youngstown, OH: *Graf Implements.*

Graham, J. E., Karmarkar, A. M., & Ottenbacher, K. J. (2012). Small sample research designs for evidence-based rehabilitation: issues and methods. *Archives of Physical Medicine and Rehabilitation*, *93*(8), 111-116. doi: 10.1016/j.apmr.2011.12.017

Greene, I., Mc Tiernan, A., & Holloway, J. (2018). Cross-Age Peer Tutoring and Fluency-Based Instruction to Achieve Fluency with Mathematics Computation Skills: A Randomized Controlled Trial. *Journal of Behavioral Education*, *27*(2), 145-171. doi: 10.1007/s10864-018-9291-1

Gregory, A., McLaughlin, T. F., Weber, K. P., & Stookey, S. (2005). The Effects of Using Direct Instruction and a Re-Reading Contingency with a High School Student. *International Journal of Special Education*, *20*(1), 50-54.

Hammersley, M. (2005). The myth of research-based practice: The critical case of educational inquiry. *International Journal of Social Research Methodology, 8*(4), 317–330. doi: 10.1080/1364557042000232844

Hartnedy, S. L., Mozzoni, M. P., & Fahoum, Y. (2005). The effect of fluency training on math and reading skills in neuropsychiatric diagnosis children: A multiple baseline design. *Behavioral Interventions*, *20*(1), 27-36. doi: 10.1002/bin/167

Haughton, E. C. (1980). Practicing practices: Learning by activity. *Journal of Precision Teaching, 1*(3), 3-20.

Head, B. W. (2015). Toward more “evidence-informed” policy making?. *Public Administration Review, 76*(3), 472-484. doi: 10.1111/puar.12475

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood programme boosted adult outcomes. *American Economic Review 2013, 103*(6). 2052-2086. doi: 10.1257/aer.103.6.2052

Henry, K. L., Knight, K. E., & Thornberry, T. P. (2012). School disengagement as a predictor of dropout, delinquency, and problem substance use during early adolescence and early adulthood. *Journal of Youth and Adolescence, 41*(2), 156-166. doi: 10.1007/s10964-011-9665-3

Higgins, S., & Katsipataki, M. (2016). Communicating comparative findings from meta-analysis in educational research: Some examples and suggestions. *International Journal of Research & Method in Education, 39*(3), 237-254. doi: 10.1080/1743727x.2016.1166486

House of Commons (2014). *Business, Innovation and Skills Committee: Adult Literacy and Numeracy*. Retrieved from https://publications.parliament.uk/pa/cm201415/cmselect/cmbis/557/557.pdf [Accessed 10 November 2018]

House, E., Glass, G., McLean, L., & Walker, D. (1978). No simple answer: Critique of the Follow Through evaluation. *Harvard Educational Review*, *48*(2), 128-160.

Hutchinson, S. & Wilson, H.S. (1992). Validity threats in scheduled semistructured research interviews. *Nursing research, 41*(2), 117-119.

Hughes, J. C., Beverley, M., & Whitehead, J. (2007). Using precision teaching to increase the fluency of word reading with problem readers. *European Journal of Behavior Analysis*, *8*(2), 221-238. doi: 10.1080/15021149.2007.11434284

Hulson-Jones, A., Hughes, J. C., Hastings, R. P., & Beverley, M. (2013). Using the toolbox series for literacy with adult struggling readers. *European Journal of Behavior Analysis, 14*(2), 349-359. doi: 10.1080/15021149.2013.11434466

Huitt, W., Huitt, M., Monetti, D., & Hummel, J. (2009). A systems-based synthesis of research related to improving students' academic performance. Paper presented at the 3rd International City Break Conference sponsored by the Athens Institute for Education and Research (ATINER), October 16-19, Athens, Greece. Retrieved from http://www.edpsycinteractive.org/papers/improving-school-achievement.pdf [last accessed 25 May 2020]

Hunter, S. H., Beverley, M., Parkinson, J., & Hughes, J. C. (2016). Increasing high school students' maths skills with the use of SAFMEDS class-wide. *European Journal of Behavior Analysis*, *17*(2), 154-165. doi: 10.1080/15021149.2016.1247643

Hunter, S. H., Geary, K., & Hughes, J. C. (2018) The effectiveness of SAFMEDS to help students learn psychology terms in high school. Presented at Association for Behavior Analysis International 44th Annual Convention, San Diego.

Jahoda, A., Hastings, R., Hatton, C., Cooper, S. A., Dagnan, D., Zhang, R., ... & Scott, K. (2017). Comparison of behavioural activation with guided self-help for treatment of depression in adults with intellectual disabilities: a randomised controlled trial. *The Lancet Psychiatry*, *4*(12), 909-919. doi: 10.1016/S2215-0366(17)30426-1

Jerrim, J., & Wyness, G. (2016). Benchmarking London in the Pisa rankings. *London Review of Education, 14*(2), 38-65. doi: 10.18546/LRE.14.2.04

Johnson, K. (1997). Morningside Academy. *Behavior and Social Issues, 7*(1), 31-35.

Johnson, K. R., & Layng, T. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American Psychologist, 47*(11), 1475-1490. doi: 10.1037/0003-066X.47.11.147

Johnson, K. R., & Layng, T. J. (1996). On terms and procedures: Fluency. *The Behavior Analyst*, *19*(2), 281-288. doi:10.1007/BF03393170

Johnson, K., & Street, E. M. (2004). *The Morningside Model of Generative Instruction: What it Means to Leave No Child Behind.* Concord, MA: Cambridge Centre for Behavioral Studies

Johnson, K., & Street, E. M. (2012). *Response to intervention and precision teaching: creating synergy in the classroom.* New York, NY: Guilford Press.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development, 74*, 834-850. doi: 10.1111/1467-8624.00571

Jordan, N. C., & Levine, S. C. (2009) Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Reviews, 15*(1), 60-68. doi: 10.1002/ddr.46

Kame'enui, E. J., Carnine, D. W., Dixon, R. C., Simmons, D. C., & Coyne, M. D. (2002). *Effective Teaching Strategies That Accommodate Diverse Learners* (2nd Ed). Upper Saddle River, New Jersey: Pearson.

Kamps, D. M., Wills, H. P., Greenwood, C. R., Thorne, S., Lazo, J. F., Crockett, J. L., ... & Swaggart, B. L. (2004). Curriculum Influences on Growth in Early Reading Fluency for Students with Academic and Behavioral Risks: A Descriptive Study. *Journal of Direct Instruction*, *4*(2), 189-210. doi: 10.1177/10634266030110040301

Katzir, T., Lesaux, N. K., & Kim, Y. (2009). The role of reading self-concept and home literacy practices in fourth grade reading comprehension. *Reading and Writing, 22*, 261-276. doi: 10.1007/s11145-007-9112-8

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological methods*, *17*(2), 137-152. doi: 10.1037/a0028086

Kerr, K. P., Smyth, P., & McDowell, C. (2003). Precision teaching children with Autism: helping design effective programmes. *Early Development and Care, 173*(4), 399-410. doi: 10.1080/0300443032000079087

Kinder, D., & Carnine, D. (1991). Direct instruction: What it is and what it is becoming. *Journal of Behavioral Education, 1*(2), 193-213. doi: 10.1007/BF00957004

Kinder, D., Kubina, R., &Marchand-Martella, N. E. (2005). Special education and direct instruction: An effective combination. *Journal of Direct Instruction, 5*(1), 1-36.

Klingner, J. K., Boardman, A. G., & McMaster, K. L. (2013). What does it take to scale up and sustain evidence-based practices? *Exceptional Children, 79*(2), 195-211. doi: 10.1177/001440291307900205

Kortesluoma, R. L., Hentinen, M., & Nikkonen, M. (2003). Conducting a qualitative child interview: methodological considerations. *Journal of Advanced Nursing*, *42*(5), 434-441.

Koutsouris, G., & Norwich, B. (2018). What exactly do RCT findings tell us in education research? *British Educational Research Journal, 44*(6), 939-959. doi: 10.1002/berj.3464

Kozioff, M. A., LaNunziata, L., Cowardin, J., & Bessellieu, F. B. (2000). Direct instruction: Its contributions to high school achievement. *The High School Journal, 84*(2), 54-71.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241-253. doi: 10.3102/0013189X20912798

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547-588. doi: 10.3102/0034654318759268

Kubina, R. M., Commons, M. L., & Heckard, B. (2009). Using precision teaching with direct instruction in a summer school programme. *Journal of Direct Instruction*, *9*(1), 1-12.

Kubina, R. M., Ward, M. C., & Mozzoni, M. P. (2000). Helping one person at a time: precision teaching and traumatic brain injury rehabilitation. *Behavioral Interventions, 15*(3), 189-203. doi: 10.1002/1099-078X(200007/09)15:3<189::AID-BIN55>3.0.CO;2-X

Kubina, R. M., & Wolfe, P. (2005). Potential applications of behavioral fluency for children with autism. *Exceptionality, 13*(1), 35-44. doi: 10.1207/s15327035ex1201_5

Kubina, R. M., & Yurich, K. K. L. (2012). *The Precision Teaching Book.* Lemont, PA: Greatness Achieved.

Lambe, D., Murphy, C., & Kelly, M. E. (2015). The impact of a precision teaching intervention on the reading fluency of typically developing children. *Behavioral interventions*, *30*(4), 364-377. doi: 10.1002/bin.1418

Lee, O., Penfield, R., & Maerten-Rivera, J. (2009). Effects of fidelity of implementation on science achievement gains among English language learners. *Journal of Research in Science Teaching*, 46(7), 836–859. doi: 10.1002/tea.20335

Leonard, S., & Woolcott, G. (2018). Contemporary issues in teaching and learning science. In G. Woolcott, & R. Whannell. (1st ed.). *Teaching Secondary Science Theory & Practice.* Singapore: Cambridge University Press

Liddy, C., Johnston, S., Irving, H., Nash, K., & Ward, N. (2015). Improving awareness, accountability, and access through health coaching: qualitative study of patients' perspectives. *Canadian Family Physician*, *61*(3), 158-164.

Lindsley, O. R. (1972). From Skinner to precision teaching: The child knows best. In Jordan, J. B., & Robbins (Eds.), *Let's try doing something else kind of thing,* 1-11. Arlington, VA: Council on Exceptional Children.

Lindsley, O. R. (1990). Precision teaching: By teachers for children. *Teaching Exceptional Children, 22*(3), 10-15. doi: 10.1177/004005999002200302

Lindsley, O. R. (1991a). Precision teaching's unique legacy from BF Skinner. *Journal of Behavioral Education, 1*(2), 253-266.

Lindsley, O. R. (1991b). BF Skinner (1904-1990): Thankyou, grandpa fred!. *Journal of Precision Teaching, 8*, 5-11.

Lindsley, O. R. (1991c). From technical jargon to plan English for application. *Journal of Applied Behavior Analysis, 24*(3), 449-458.

Lindsley, O. R. (1992). Precision teaching: Discoveries and effects. *Journal of Applied Behavior Analysis*, *25*(1), 51-57.

Lindsley, O. R. (1995). Precision Teaching: By teachers for children. *Journal of Precision Teaching, 12*(2), 9-17.

Lindsley, O. R. (1996a). Is fluency free-operant response-response chaining?. *The Behavior Analyst, 19*(2), 211-224. doi: 10.1007%2FBF03393165

Lindsley, O. R. (1996b). The four free-operant freedoms. *The Behavior Analyst, 19*(2), 199-210. doi: 10.1007/BF03393164

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., ... & Busick, M. D. (2012). Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms. *National Center for Special Education Research*. Retrieved online https://files.eric.ed.gov/fulltext/ED537446.pdf [Accessed 20 May 2020]

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher,* 1-9. doi: 10.3102/00313189X19832850

Lortie-Forgues, H., Tian, J., & Siegler, R. S. (2015). Why is learning fraction and decimal arithmetic so difficult?. *Developmental Review*, *38*, 201-221. doi: 10.1016/j.dr.2015.07.008

Lynch, K., Hill, H.C., Gonzalez, K, & Pollard, C. (2019). Strengthening the Research Base that Informs STEM Instructional Improvement Efforts: A Meta-Analysis. *Education Evaluation and Policy Analysis, 41*(3), 260-293. doi: 10.3102/0162373719849044

Mason, L. L., Rivera, C. J., & Arriaga, A. (2018). The effects of an avoidance contingency on post-secondary student SAFMEDS performance. *European Journal of Behavior Analysis, 19*(1), 62-71. doi: 10.1080/15021149.2017.1418125

May, T., Rinehart, N., Wilding, J., & Cornish, K. (2013). The role of attention in the academic attainment of children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*(9), 2147-2158. doi: 10.1007/s10803-013-1766-2

McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, G., Fisher, P. (2007)., The Hawthorne effect: a randomised, controlled trial. *BMC Medical Research Methodology, 7*(30). doi: 10.1186/1471-2288-7-30

McDowell, C., & Keenan, M. (2001). Developing fluency and endurance in a child diagnosed with attention deficit hyperactivity disorder. *Journal of Applied Behavior Analysis*, *34*(3), 345-348. doi: 10.1901/jaba.2001.34-345.

McGraw-Hill (2019). *Corrective Mathematics.* Retrieved from https://www.mheducation.com.au/schools/numeracy/corrective-mathematics [Accessed 26 May 2020]

McGreevy, P. (1983). *Teaching and learning in plain English* (2nd ed.). Kansas City: Plain English Publications.

Meindl, J. N., Ivy, J. W., Miller, N., Neef, N. A., & Williamson, R. L. (2013). An examination of stimulus control in fluency-based strategies: SAFMEDS and generalisation. *Journal of Behavioral Education, 22*(3), 229-252. doi: 10.1007/s10864-013-9172-6

Merbitz, C., Vieitez, D., Hansen Merbitz, N. & Binder, C. (2004). Precision Teaching: Applications in Education and Beyond. In: D. Moran and R. Malott, ed., *Evidence-Based Educational Methods* (Eds). San Diego, California: Elsevier Academic Press, pp.63-78.

Merbitz, C., Vieitez, D., Merbitz, N. H., & Pennypacker, H. S. (2004). Precision teaching: Foundations and classroom applications. In D. J., Moran, & R. W. Malott, (2004). *Evidence-Based Educational Methods* (pp. 47-62). San Diego, California: Elsevier Academic Press.

Miller, A. D., & Heward, W. L. (1992). Do your students really know their math facts? Using daily time trials to build fluency. *Intervention in School and Clinic, 28*(2), 98-104.

Missall, K. N., Mercer, S. H., Martínez, R. S., & Casebeer, D. (2012). Concurrent and longitudinal patterns and trends in performance on early numeracy curriculum-based measures in kindergarten through third grade. *Assessment for Effective Intervention, 37*(2), 95-106. doi: 10.1177/1534508411430322

Moran, D. J., & Malott, R. W. (2004). *Evidence-Based Educational Methods.* San Diego, California: Elsevier Academic Press.

Morrell, M., Morrell, R., & Kubina, R. M. (1995). Using precision teaching to enhance direct instruction sight word reading. *Journal of Precision Teaching and Celeration, 13*(1), 47-54.

Morrison, K. (2001). Randomised controlled trials for evidence-based education: some problems in judging ‘what works’. *Evaluation & Research in Education, 15*(2), 69-83. doi: 10.1080/09500790108666984

Munizi, A., & Mc Laughlin, T. F. (2013). The effects of an error drill procedure on the sight-word reading skills of an elementary student with severe behavior disorders. *International Journal of English Education, 2*(3), 493-500.

Murray, D. M. & Blitstein, J. L. (2003). Methods to reduce the impact of intraclass correlation in group-randomized trials. *Evaluation Review, 27*(1), 79-103. doi: 10.1177/0193841X02239019

Nam, S. S. & Spruill, M. (2005). Learning channel intervention to develop and generalize fluency in multiplication facts. *Journal of Early and Intensive Behavior Intervention, 2*(2), 103-111. doi: 10.1037/h0100305

Nelson, J., & Campbell, C. (2017). Evidence-informed practice in education: meanings and applications. *Educational Research, 59*(2), 127-135. doi: 10.1080/00131881.2017.1314115

Nelson, P. M., Burns, M. K. Kanive, R., & Ysseldyke, J. E. (2013). Comparison of math fact rehearsal and a mnemonic strategy approach for improving math fact fluency. *Journal of School Psychology, 51*, 659-667. doi: 10.1016/j.jsp.2013.08.003

Nunes, T., Bryant, P., Sylva, K., & Barros, R. (2009). *Development of maths capabilities and confidence in primary schools*. Retrieved from http://dera.ioe.ac.uk/11154/1/DCSF-RR118.pdf [Accessed: 17 September 2018].

O'Connell, S., & SanGiovanni, J. (2011). *Mastering the basic math facts in addition and subtraction. Portsmouth, NH: Heinemann*

OECD (2010). PISA 2009 results: What students know and can do – student performance in reading, mathematics, and science (volume I). doi: 10.1987/9789264091450-en

OECD (2011). *Against the odds: Disadvantaged students who succeed in school.* Paris: OECD Publishing

OECD (2012), Equity and Quality in Education: Supporting Disadvantaged Students and Schools, OECD Publishing, Paris. doi: 10.1787/9789264130852-en.

OECD (2016a). PISA 2015 results (Volume I): Excellence and equity in education. Paris: OECD Publishing. doi: 10.1787/9789264266490-en

OECD (2016b). *What makes a school a learning organisation? A guide for policy makers, school leaders and teachers.* Retrieved from https://www.oecd.org/education/school/school-learning-organisation.pdf [Accessed 25 August 2020]

Ofsted (2016). Alternative provision: *The findings from Ofsted's three-year survey of schools' use of off-site alternative provision.* Retrieved from https://www.gov.uk/government/publications/alternative-school-provision-findings-of-a-three-year-survey [Accessed 7 September 2020]

Ojose, B. (2011). Mathematics literacy: Are we able to put the mathematics we learn into everyday use. *Journal of Mathematics Education, 4*(1), 89-100.

Owen, K. O., Hunter, S. H., Watkins, R., Payne, J., Bailey, T., Hastings, R., & Hughes, J. C. (under review). *A Randomized Controlled Trial of Precision Teaching Support During a Fluency-Based Numeracy Intervention.* Manuscript submitted for publication.

Parliament UK (2014). *Underachievement in education by white working class children - Education committee.* Retrieved from https://www.publications.parliament.uk/pa/cm201415/cmselect/cmeduc/142/14205.htm [Accessed 14 October 2018]

Parliament UK (2018). *Forgotten children: Alternative provision and the scandal of ever increasing exclusions*. Retrieved from https://publications.parliament.uk/pa/cm201719/cmselect/cmeduc/342/34208.htm#footnote-020 [Accessed 17 January 2020]

Parsons, J. L., Marchand-Martella, N. E., Waldron-Soler, K., Martella, R. C., & Lignugaris/Kraft, B. (2004). Effects of a high school-based peer-delivered *Corrective Mathematics* programme. *Journal of Direct Instruction, 4*(1), 95–103.

Passolunghi, M. C., & Pazzaglia, F. (2005). A comparison of updating processes in children good or poor in arithmetic word problem-solving. *Learning and Individual Differences, 15*(4), 257-269. doi: 10.1016/j.lindif.2005.03.001

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. California: Sage Publications.

Perry, C. (2016). *Research and Information Service: Briefing Note*. Northern Ireland Assembly. Retrieved from http://www.niassembly.gov.uk/globalassets/documents/raise/publications/2016-2021/2016/education/5216.pdf [Accessed 16 October 2018]

Peterson, S. K., Hudson, P. J., Mercer, C. D., & McLeod, P. D. (1990). The effect of the concrete to abstract teaching sequence on acquisition and retention of place value skills. *Journal of Precision Teaching and Celeration, 7*, 79-84

Piaggio, G., Carroli, G., Villar, J., Pinol, A., Bakketeig, L., Lumbiganon, P., ... & Farnot, U. (2001). Methodological considerations on the design and analysis of an equivalence stratified cluster randomization trial. *Statistics in medicine*, *20*(3), 401-416. doi: 10.1002/1097-0258(20010215)20:3<401::AID-SIM801>3.0.CO;2-1

Potts, L., Eshleman, J. W., & Cooper, J. O. (1993). Ogden R. Lindsley and the historical development of precision teaching. *The Behavior Analyst, 16*, 177-189. doi: 10.1007/BF03392622

Pryzychodzin, A. M., Marchand-Martella, N. E., Martella, R. C., & Azim, D. (2004). Direct instruction mathematics programmes: An overview and research summary. *Journal of Direct Instruction, 4*(1), 53-84.

Przychodzin-Havis, A. M., Marchand-Martella, N. E., Martella, R. C., Miller, D. A., Warner, L., Leonard, B., & Chapman, S. (2005). An Analysis of" Corrective Reading" Research. *Journal of Direct Instruction*, *5*(1), 37-65.

Quigley, S. P., Peterson, S. M., Frieder, J. E., & Peck, K. M. (2018). A review of SAFMEDS: Evidence for procedures, outcomes and directions for future research. *Perspectives on Behavior Science,* 1-19. doi: 10.1007/s40614-017-0087-8

Ramey, D., Lydon, S., Healy, O., McCoy, A., Holloway, J., & Mulhern, T. (2016). A systematic review of the effectiveness of precision teaching for individuals with developmental disabilities. *Review Journal of Autism and Developmental Disorders*, *3*(3), 179-195. doi: 10.1007/s40489-016-0075-z

Reich, N. G. & Milstone, A. M. (2014). Improving efficiency in cluster-randomized study design and implementation: taking advantage of a crossover. *Open Access Journal of Clinical Trials, 6*, 11-15. doi: 10.2147/OAJCT.S56730

Resnick, L. B. and Ford, W. W. (2008). *Psychology of mathematics for instruction*. Abingdon, United Kingdom: Routledge. (Original work published in 1981)

Reyna, V. F., Nelson, W. L., Han, O. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin,135*(6), 943-973.

Reynolds, C. R., Voress, J. K., & Kamphaus, R. W. (2015). *Mathematics Fluency and Calculation Tests: Elementary.* Texas: Pro-ed

Ritchie, J. & Lewis, J. (2003). *Qualitative research practice: A guide for social science students and researchers*. London: Sage publications.

Rockwell, S. B., Griffin, C. C., & Jones, H. A. (2011). Schema-based strategy instruction in mathematics and the word problem-solving performance of a student with autism. *Focus on Autism & Other Developmental Disabilities, 26*(2), 87-95. doi: 10.1177/1088357611405039

Sailors, M., & Shanklin, N. L. (2010). Introduction: Growing evidence to support coaching in literacy and mathematics. *Elementary School Journal, 111*(1), 1-6. doi: 10.1086/653467

Saville, B. K., & Buskist, W. (2003). Traditional idiographic approaches: Small-N research designs. In Davis, S. F., Saville, B. K., & Buskist, W. (Eds.), *Handbook of Research Methods in Experimental Psychology*, (pp. 66–82). Hoboken: Blackwell Publishing Ltd.

Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(1), 32-43. doi: 10.1007/s10488-010-0321-0

Schwartz, I. S., & Baer, D. M. (1991). Social validity assessments: Is current practice state of the art? *Journal of Applied Behavior Analysis, 24*, 189-204. doi: 10.1901/jaba.1991.24-189

Simonsen, F., & Gunter, L. (2001). Best practices in spelling instruction: A research summary. *Journal of Direct Instruction, 1*(2), 97-105.

Simpson, A. (2017). The misdirection of public policy: Comparing and combining standardised effect sizes. *Journal of Education Policy, 32*(4), 450-466. doi: 10.1080/02680939.2017.1280183

Simpson, A. (2019). Separating arguments from conclusions: the mistaken role of effect size in educational policy research. *Educational Research and Evaluation, 25*(1-2), 99-109. doi: 10.1080/13803611.2019.1617170

Slocum, T. A. (2004). Direct Instruction: The Big Ideas. In D. J., Moran, & R. W. Malott, (2004). *Evidence-Based Educational Methods* (pp. 81-94). San Diego, California: Elsevier Academic Press.

Sommers, J. (1991). Direct Instruction programs produce significant gains with at-risk middle-school students. *Direct Instruction News, 11*(1), 7–14.

Stockard, J., Wood, T. W., Coughlin, C., & Khoury, C. R. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research, 88*(4), 479-507. doi 10.3102/0034654317751919

Stockwell, F., & Eshelman, J. (2010). A case-study using SAFMEDS to promote fluency with Skinner's verbal behaviour terms. *Journal of Precision Teaching and Celeration, 26*, 33-40.

Street, E. M., & Johnson, K. (2014). The sciences of learning instruction and assessment as underpinnings of the Morningside model of generative instruction. *Acta de investigaciógica, 4*(3), 1772-1792. doi: 10.1016/S2007-4719(14)70979-2

Strømgren, B., Berg-Mortensen, C., & Tangen, T. (2014). *European Journal of Behavior Analysis, 15*(2), 225-240. doi: 10.1080/15021149/2014.11434723

Swanson, H. L. (1999). Instructional components that predict treatment outcomes for students with learning disabilities: Support for a combined strategy and direct instruction model. *Learning Disabilities Research & Practice, 14*(3), 129-140. doi: 10.1207/sldrp1403_1

Taylor, C. (2017). The reliability of free school meal eligibility as a measure of socio-economic disadvantage: Evidence from the millennium cohort study in Wales. *British Journal of Educational Studies*, 66(1), 1-23. doi: 10.1080/00071005.2017.1330464

Thompson, J. L., Wood, C. L., Test, D. W., & Cases-Cook, J. (2012). Effects of direct instruction on telling the time by students with autism. *Journal of Direct Instruction, 12*, 1-12.

Thornicroft. G. H., Lempp., H., & Tansella, M. (2011). The place of implementation science in the translational medicine continuum. *Psychological Medicine, 41*(11), 2015-2021. doi: 10.107/S0033291711000109

Tincani, M. (2004). Improving outcomes for college students with disabilities: Ten strategies for instructors. *College Teaching, 52*(4), 128-132. doi: 10.3200/CTCH.52.4.128-133

Titeca, D., Roeyers, H., Josephy, H., Ceulemans, A., & Desoete, A. (2014). Predictors of mathematics in first grade children with autism spectrum disorder. *Research in Developmental Disabilities, 35*(11), 2714-2727. doi: 10.1016/j.ridd.2014.07.012

Tolich, M., & Davidson, C. (2003). Collecting the data. In Davidson, C., & Tolich, M. (Eds), *Social science research in New Zealand* (2nd ed., pp. 121-153). Auckland: Pearson Education.

Tuval-Mashiach, R. (2017). Raising the curtain: The importance of transparency in qualitative research. *Qualitative Psychology*, *4*(2), 126-138. doi: 10.1037/QUP0000062

Tyler, E., Watkins, R. C., Roberts, S. E., Hoerger, M., Hulson-Jones, A., Hughes, J. C., & Hastings, R. P. (2019). Research note: Collaborative Institute for Education Research, Evidence and Impact (CIEREI). *Wales Journal of Education, 20*(1), 135-140. doi: 10.16922/wje.20.1.8

Van Houten, R. (1984). Setting up performance feedback systems in the classroom. In Heward, W. L., Heron, T. E., Hill, D. S., & Trap-Porter, J. (Eds), *Focus on behavior analysis in education* (pp.114-125). Columbus, OH: Merrill

Van Rinsveld, A., Dricot, L., Guillaume, M., Rossion, B., & Schiltz, C. (2017). Mental arithmetic in the bilingual brain: Language matters. *Neuropsychologia*, *101*, 17-29. doi: 10.1016/j.neuropsychologia.2017.05.009

Van Selm, M., & Jankowski, N. W. (2006). Conducting online surveys. *Quality and Quantity*, *40*(3), 435-456. doi: 10.1007/s11135-005-8081-8

Vargas, J. S. (2013). *Behavior Analysis for Effective Teaching*. Routledge.

Watkins, C. L. (1997). *Project Follow Through: A case study of contingencies influency instructional practices of the educational establishment.* Cambridge Massachusetts: Cambridge Center for Behavioral Studies.

Welsh Government (2013a). *National Literacy and Numeracy Framework*. Retrieved from https://hwb.gov.wales/storage/64812b0e-3d20-4d10-94e8-68221e47a38d/to-support-schools-in-introducing-the-national-literacy-and-numeracy-framework.pdf [Accessed 17 December 2019]

Welsh Government (2013b). *Guidance for using the pupil deprivation grant: What really works?*. Retrieved from https://gov.wales/sites/default/files/publications/2018-11/what-really-works.pdf [Accessed 7 September 2020]

Welsh Government (2015). *Curriculum for Wales: Foundation Phase Framework (revised 2015)*. Retrieved from http://learning.gov.wales/docs/learningwales/publications/150803-fp-framework-en.pdf [Accessed 25 November 2019]

Welsh Government (2016). *Programme of Study for Mathematics Key Stages 2-4*. Retrieved from https://hwb.gov.wales/storage/e2aad4fc-faae-4054-a382-242d08ee2021/programme-of-study-for-mathematics.pdf [Accessed 17 December 2019]

Welsh Government (2018). *A handbook for Management Committees of Pupil Referral Units.* Retrieved from https://gov.wales/sites/default/files/publications/2018-11/handbook-for-management-committees-of-pupil-referral-units.pdf [Accessed 21 June 2019]

Welsh Government (2019). *Draft statutory guidance area of learning and experience: Mathematics and numeracy*. Retrieved from https://hwb.gov.wales/storage/ee62ba8e-b15d-4dda-aad5-b8ba05100be6/draft-statutory-guidance-mathematics-and-numeracy.pdf [Accessed 26 November 2019]

Westfall, J. M., Mold, J., & Fagnan, L. (2007). Practice-based research- 'Blue highways' on the NIH roadmap. *Journal of the American Medical Association, 297*(4), 403-406. doi: 10.1001/jama.297.4.403

White, O. R. (1986). Precision teaching—precision learning. *Exceptional children*, *52*(6), 522-534. doi: 10.1177/001440298605200605

White, O. R. (1986). Precision teaching—precision learning. *Exceptional children*, *52*(6), 522-534.

White, O, R. (2000). *Decisions, Decisions* ... Retrieved online https://education.uw.edu/sites/default/files/areas/edspe/white/docs/Decisions_Decisions.pdf [Accessed 27 July 2018]

White, O. R., & Neely, M. D. (2012). *The Chart Book: An overview of standard celeration chart conventions and practices.* Retrieved from https://education.uw.edu/sites/default/files/areas/edspe/white/readings/chartbook.pdf [Accessed 1 July 2019]

Willig, C. (2008). *Introducing Qualitative Research in Psychology.* McGraw-Hill Education.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test, Fourth Edition.* Lutz, Florida: Pearson Education Ltd [measurement instrument].

Wolery, M., Ault, M. J., Doyle, P. M., Gast, D. L., & Griffen, A. K. (1992). Choral and individual responding during small group instruction: Identification of interactional effects. *Education and Treatment of Children*, *15*(4), 289-309.

Wolf, A. (2011). *Review of vocational education.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/180504/DFE-00031-2011.pdf [Accessed 15 November 2019]

Wyse, D. (2001). Promising yourself to do better?: Target-setting and literacy. *Education 3-13, 29*(2), 13-18. doi: 10.1080/03004270185200161

Wyse, D., Brown, C., Oliver, S., & Poblete, X. (2018). *The BERA close-to-practice research project: Research report*. Retrieved from https://www.bera.ac.uk/publication/bera-statement-on-close-to-practice-research [Accessed 29 July 2020]

XLP (n.d.a). *Homepage.* Retrieved from https://www.xlp.org.uk [Accessed 09 March 2020]

XLP (n.d.b). *Vision and Values.* Retrieved from https://www.xlp.org.uk/vision-and-values.html [Accessed 20 March 2020]

XLP (n.d.c). *What we do: Education*. Retrieved from http://www.xlp.org.uk/what-we-do/education [Accessed 15 July 2018]

Ysseldyke, J., Spicuzza, R., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance mathematics achievement in Urban schools. *Journal of Education for Students Placed at Risk, 8*(2), 247-265. doi: 10.1207/S15327671ESPR0802_4

Zimmermann, C., Swami, N., Krzyzanowska, M., Hannon, B., Leighl, N., Oza, A., ... & Donner, A. (2014). Early palliative care for patients with advanced cancer: a cluster-randomised controlled trial. *The Lancet*, *383*(9930), 1721-1730. doi: 10.1016/S0140-6736(13)62416-2

Appendices

Appendix A: Sensitivity analysis (Chapter 3)

Table A1

Summary of the model outcomes based on the dataset containing missing data.

	On-going support		No support		Adjusted cases analysis*				
	n	Marginal mean of raw scores [95% CI]	*n*	Marginal mean of raw scores [95% CI]	Adjusted difference between change scores [95% *CI*]	*p*	*d* [95% CI]		ICC
MFaCTs: Grades 1-2									
Baseline	285	12.00 [10.28, 13.73]	269	9.02 [7.28, 10.76]	-	-	-	School	0.36
Follow-up	256	22.59 [20.85, 24.34]	237	16.50 [14.72, 18.29]	2.92 [1.22, 4.61]	.001	0.23 [0.06, 0.40]	Children	0.46
MFaCTs: Grades 3-5									
Baseline	285	8.52 [7.00, 10.05]	269	5.94 [4.40, 7.48]	-	-	-	School	0.37
Follow-up	251	19.14 [17.59, 20.69]	238	13.76 [12.19, 15.34]	2.68 [1.21, 4.14]	<.001	0.25 [0.08, 0.42]	Children	0.42

* Model adjusted for the following fixed effects: county, child's predominant language, gender, free school meals status, school year group. Interaction effect: trial arm x time. Random effects: school, children.

Table A2

Summary of The Model Outcomes Based on a Complete Case Analysis

	On-going support		No support		Adjusted cases analysis*				
	n	Marginal mean of raw scores [95% CI]	*n*	Marginal mean of raw scores [95% CI]	Adjusted difference between change scores [95% *CI*]	*p*	*d* [95% CI]		ICC
MFaCTs: Grades 1-2									
Baseline	248	11.53 [9.71, 13.35]	229	8.77 [6.90, 10.65]	-	-	-	School	0.37
Follow-up	248	22.40 [20.58, 24.22]	229	16.33 [14.46, 18.21]	3.32 [1.60, 5.04]	<.001	0.26 [0.08, 0.44]	Children	0.46
MFaCTs: Grades 3-5									
Baseline	243	8.22 [6.61, 9.82]	230	5.63 [4.00, 7.27]	-	-	-	School	0.38
Follow-up	243	18.95 [17.34, 20.55]	230	13.60 [11.96, 15.24]	2.76 [1.25, 4.27]	<.001	0.26 [0.08, 0.44]	Children	0.42

* Model adjusted for the following fixed effects: county, child's predominant language, gender, free school meals status, school year group. Interaction effect: trial arm x time. Random effects: school, children.

Table A3

Summary of The Model Outcomes Based on 50 Iterations of Imputed Data

	On-going support		No support		Adjusted cases analysis*				
	n	Marginal mean of raw scores [95% CI]	*n*	Marginal mean of raw scores [95% CI]	Adjusted difference between change scores [95% *CI*]	*p*	*d* [95% CI]		ICC
MFaCTs: Grades 1-2									
Baseline	294	12.38 [10.71, 14.05]	281	9.05 [7.38, 10.72]	-	-	-	School	0.32
Follow-up	294	22.62 [20.87, 24.36]	281	16.62 [14.87, 18.37]	2.67 [0.75, 4.58]	.006	0.21 [0.04, 0.37]	Children	0.43
MFaCTs: Grades 3-5									
Baseline	294	8.70 [7.25, 10.15]	281	6.00 [4.54, 7.47]	-	-	-	School	0.33
Follow-up	294	19.18 [17.70, 20.67]	281	13.66 [12.15, 15.18]	2.81 [1.20, 4.42]	.001	0.26 [0.09, 0.42]	Children	0.40

* Model adjusted for the following fixed effects: county, child’s predominant language, gender, free school meals status, school year group. Interaction effect: trial arm x time. Random effects: school, children

Table A4

Baseline measurements for the children's outcome measures, based on the dataset containing missing data

	No Support			Support		
	n	*M (SD)*	Min to max	*n*	*M (SD)*	Min to max
Measure						
MFaCTs: Grades 1-2	281	9.06 (6.94)	0 to 36.01	294	12.52 (10.52)	0 to 65.28
MFaCTs: Grades 3-5	281	5.99 (4.89)	0 to 26.76	294	8.85 (8.01)	0 to 44.21

Table A5

Baseline measurements for the children's outcome measures, based on complete case analysis

	No Support			Support		
	n	*M (SD)*	Min to max	*n*	*M (SD)*	Min to max
Measure						
MFaCTs: Grades 1-2	229	8.87 (6.82)	0 to 35	248	11.24 (11.12)	0 to 60
MFaCTs: Grades 3-5	230	5.71 (4.71)	0 to 26	243	7.88 (6.63)	0 to 35

Table A6

Baseline measurements for the children's outcome measures, based on imputed dataset

	No Support			Support		
	n	*M (SD)*	Min to max	*n*	*M (SD)*	Min to max
Measure						
MFaCTs: Grades 1-2	281	9.06 (6.94)	0 to 36.01	294	12.52 (10.52)	0 to 65.28
MFaCTs: Grades 3-5	281	5.99 (4.89)	0 to 26.76	294	8.85 (8.01)	0 to 44.21

Appendix B: Moderation analyses (Chapter 3)

Table A7

Outcomes for moderation analysis between gender, trial arm, and time

	MFaCTs: Grades 1-2		**MFaCTs: Grades 3-5**	
Interaction	Adjusted difference between change scores * [95% CI]	*p*	Adjusted difference between change scores * [95% CI]	*p*
Gender x trial arm	1.44 [-1.63, 4.57]	.35	1.20 [-1.31, 3.71]	.35
Gender x time	0.01 [-2.40, 2.42]	.99	0.79 [-1.28, 2.86]	.45
Trial arm x time	2.22 [-0.17, 4.60]	.07	2.33 [0.26, 4.49]	.03
Gender x trial arm x time	1.38 [-1.99, 4.76]	.42	0.69 [-2.23, 3.60]	.64

* Model adjusted for the following fixed effects: county, child's predominant language, free school meals status, school year group. Random effects: school, children

Table A8

Outcomes for moderation analysis between a measure of social deprivation (eFSM), trial arm, and time

	MFaCTs: Grades 1-2		**MFaCTs: Grades 3-5**	
Interaction	Adjusted difference between change scores* [95% CI]	*p*	Adjusted difference between change scores* [95% CI]	*p*
eFSM x trial arm	0.36 [-3.21, 3.93]	0.84	-1.15 [-4.04, 1.74]	.43
eFSM x time	-1.18 [-3.92, 1.55]	0.40	-1.06 [-3.42, 1.30]	.38
Trial arm x time	3.27 [1.32, 5.22]	.001	2.97 [1.29, 4.66]	.001
eFSM x trial arm x time	-1.50 [-5.36, 2.37]	.45	-1.28 [-4.62, 2.07]	.45

* Model adjusted for the following fixed effects: county, child's predominant language, gender, school year group. Random effects: school, children.

Table A9

Outcomes for moderation analysis between language, trial arm, and time

	MFaCTs: Grades 1-2		**MFaCTs: Grades 3-5**	
Interaction	Adjusted difference between change scores* [95% CI]	*p*	Adjusted difference between change scores* [95% CI]	*p*
Language x trial arm	2.12 [-3.30, 7.54]	.44	1.02 [-3.70, 5.74]	.67
Language x time	-0.36 [-0.39, 2.35]	.79	-0.36 [-2.70, 1.98]	.76
Trial arm x time	1.11 [-2.19, 4.41]	.51	1.88 [-0.95, 4.72]	.19
Language x trial arm x time	2.43 [-1.42, 6.28]	.22	1.08 [-2.34, 4.39]	.52

* Model adjusted for the following fixed effects: county, gender, free school meals status, school year group. Random effects: school, children

Appendix C: Online survey questions (Chapter 4)

Please note: Teaching staff could leave any of the questions blank if they did not wish to answer them. All questions were accompanied by a free-type text box so teaching staff could provide as much or as little detail as they wanted to.

Item number	English	Cymraeg
1	Please write your name (we will only use this information to remove your data if you decide to withdraw at a later date).	Ysgrifennwch eich enw (byddwn yn defnyddio'r wybodaeth hon yn unig i gael gwared ar eich data os penderfynwch dynnu'n ôl yn ddiweddarach).
2	Which school do you teach in?	Ym mha ysgol ydych chi'n dysgu?
3	What is your job title within the school?	Beth yw teitl eich swydd yn yr ysgol?
4	Please provide us with some details about who runs the SAFMEDS sessions in your school:	Rhowch fanylion i ni ynghylch pwy sy'n cynnal y sesiynau SAFMEDS yn eich ysgol:
5	What skill(s) have you been using SAFMEDS to improve in your school (i.e., numeracy, literacy)?	Pa sgil(iau) yn eich ysgol (h.y. rhifedd, llythrennedd) ydych chi wedi bod yn defnyddio SAFMEDS i'w gwella?
6	What do you think the benefits are from using SAFMEDS?	Beth ydych chi'n ei feddwl yw'r manteision o ddefnyddio SAFMEDS?
7	Are there any factors that you think help the success of a SAFMEDS session? Please explain your answer.	A oes unrhyw ffactorau sy'n helpu i sicrhau llwyddiant sesiwn SAFMEDS yn eich barn chi? Eglurwch eich ateb.
8	Are there any factors that you think hinder the success of a SAFMEDS session? Please explain your answer.	A oes unrhyw ffactorau sy'n atal llwyddiant sesiwn SAFMEDS yn eich barn chi? Eglurwch eich ateb.
9	Have you seen a difference in the children’s behaviour as a result of the SAFMEDS sessions? Please explain your answer.	Ydych chi wedi gweld gwahaniaeth yn ymddygiad y plant o ganlyniad i'r sesiwn SAFMEDS? Eglurwch eich ateb.
10	Is there anything that the children do during a SAFMEDS session that you think hinders their performance?	Oes yna unrhyw beth mae'r plant yn ei wneud yn ystod sesiwn SAFMEDS sy'n atal eu perfformiad yn eich barn chi?
11	Would you like to use SAFMEDS in the future? Please justify your response.	Fyddech chi'n hoffi defnyddio SAFMEDS yn y dyfodol? A wnewch chi gyfiawnhau eich ymateb os gwelwch yn dda.
12	Please use the space below to provide the details of any children’s SAFMEDS performance that you would like to share.	Defnyddiwch y lle gwag isod i roi manylion am unrhyw berfformiad SAFMEDS plant yr hoffech chi ei rannu.

Appendix D: Semi-structured interview questions (Chapters 4 and 5)

Main question	Prompts
Do you enjoy using SAFMEDS?	• Why do you/don't you enjoy using it?
What is your favourite thing about SAFMEDS?	• What do you like the most about coming out of class and using SAFMEDS? • Why is that your favourite thing?
What is your least favourite thing about SAFMEDS?	• What do you like the least about coming out of class and using SAFMEDS? • If you could change one thing about coming out, what would you change? • Why is that your least favourite thing?
Do you think that SAFMEDS is useful?	• Why do you think that it is/isn't useful?
Do you use SAFMEDS outside of school?	• How? • Why?

Appendix E: Follow-up interview questions (Chapter 6)

Main question	Prompts
Do you enjoy coming out of class and doing maths with me?	• Why do you/don't you enjoy it? • What parts of the session do you/don't you enjoy? • What's your favourite part of the sessions? • What is your least favourite part of the sessions?
Do you think that coming out to do these sessions is useful?	• Why do you think that the sessions are/aren't useful?
Would you like to carry on doing these maths sessions?	• Why/why not?

Appendix F: Tom's, Will's, Leo's, and Chris' SCCs (Chapter 6)

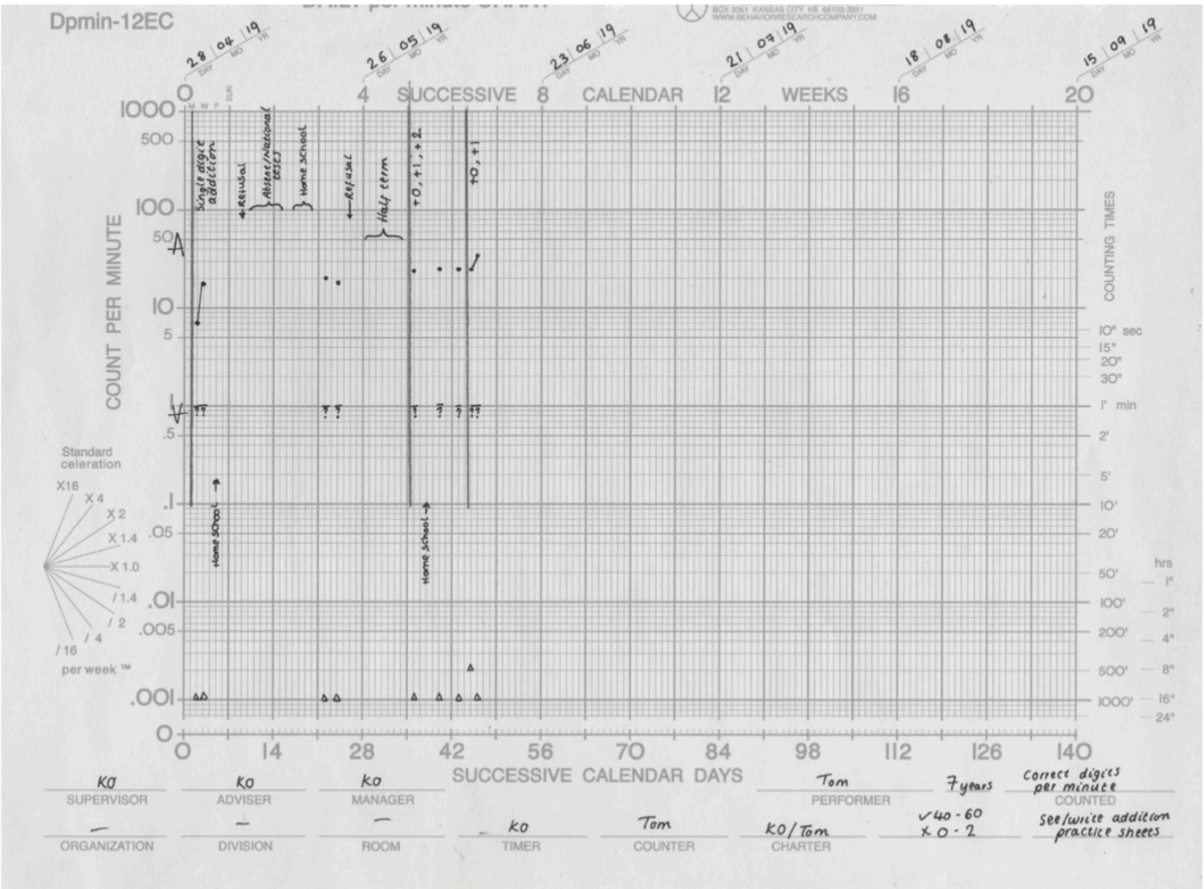

Figure A1. Tom's SCC

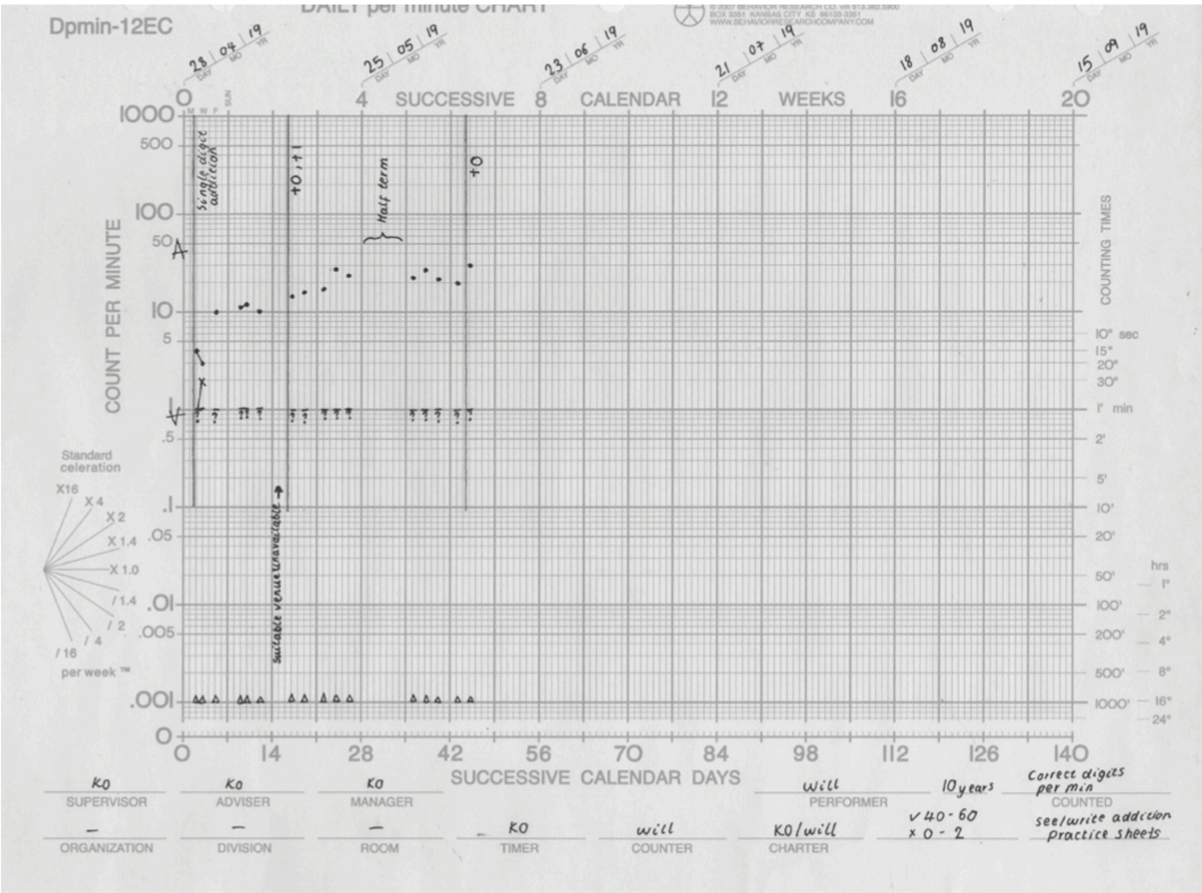

Figure A2. Will's SCC

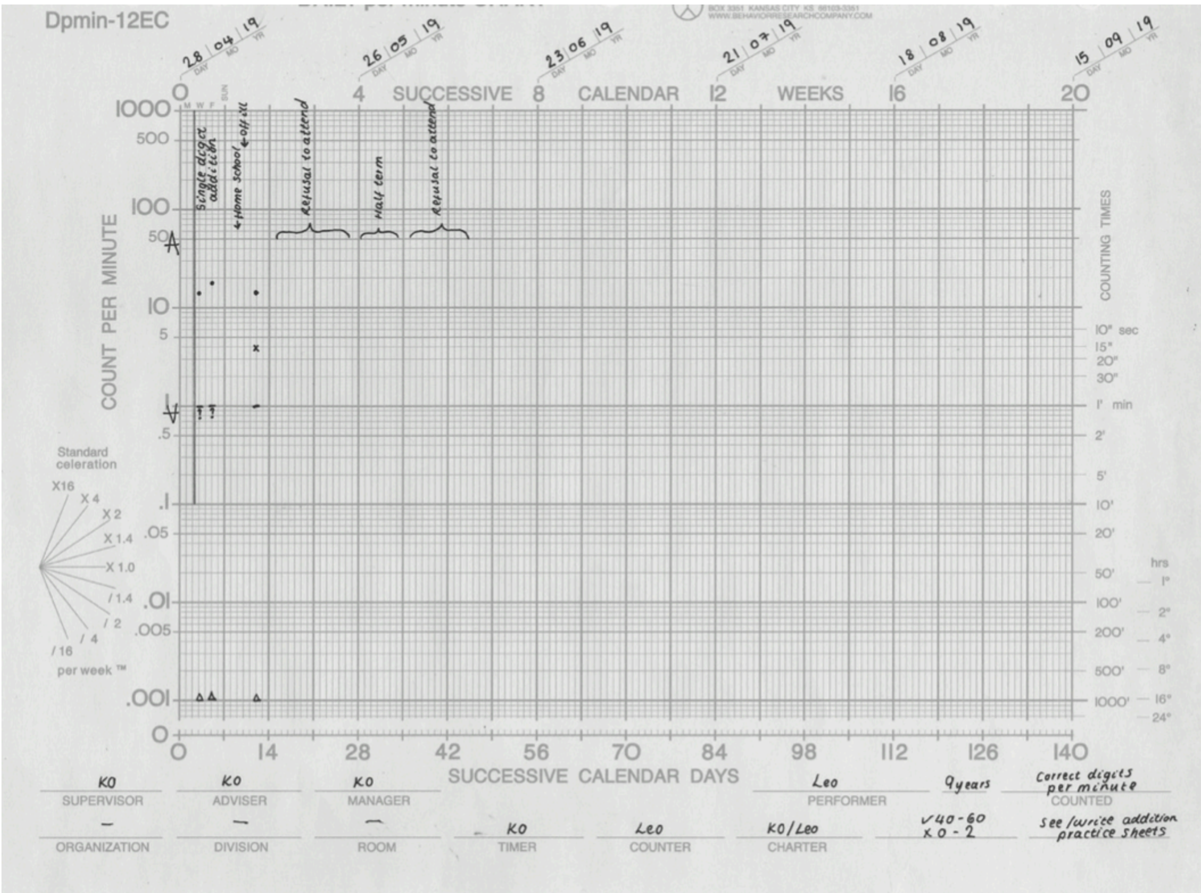

Figure A3. Leo's SCC

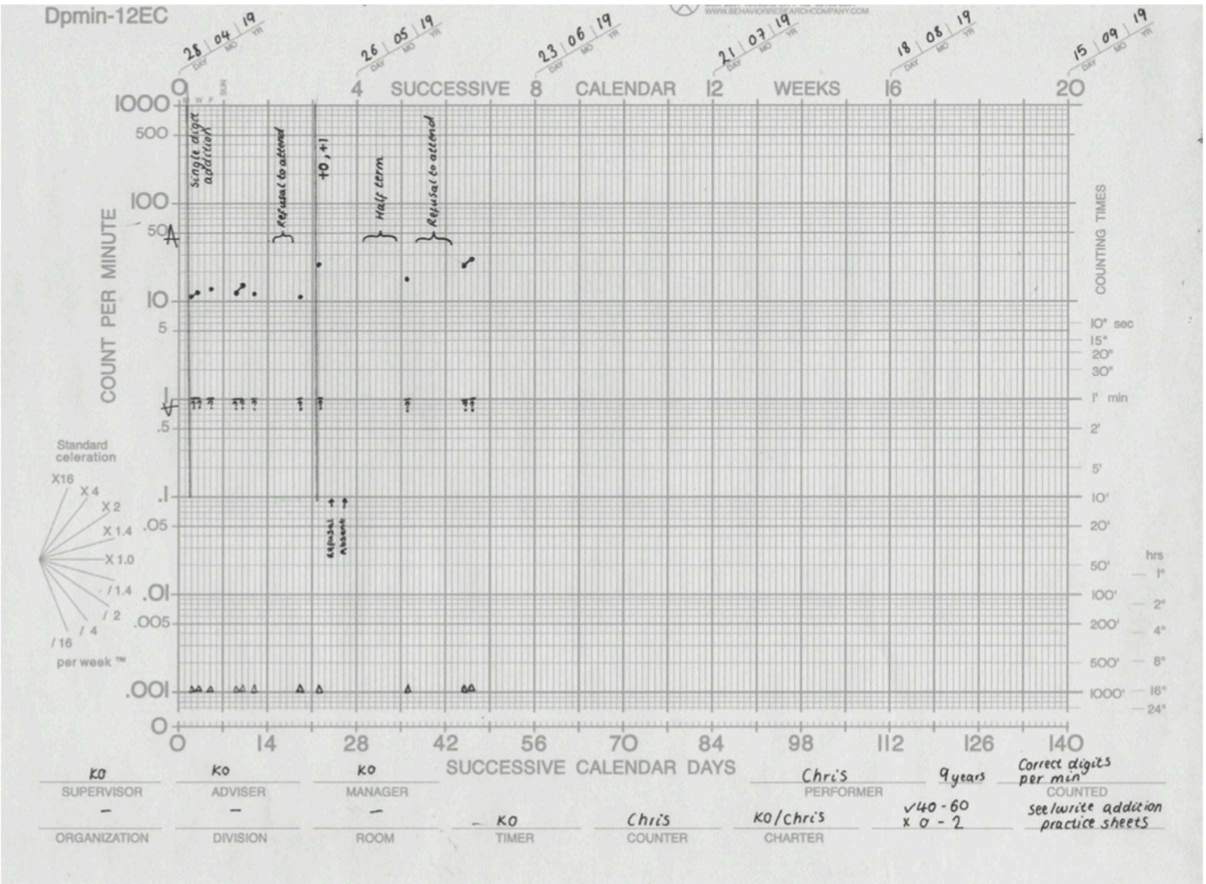

Figure A4. Chris' SCC