Do structured manipulatives aid understanding of additive reasoning, and addition and subtraction fluency in a sample of Year 7 and 8 students?

By

Daniel Green

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree

of Master of Education

Victoria University of Wellington

July, 2020

Abstract

Mathematical achievement may impact on outcomes in later life; thus, identifying and improving key mathematical skills is a focus of a large body of educational research. Both additive reasoning, and knowledge of addition and subtraction facts, appear to predict later mathematical achievement. The current study explores the impact of a short intervention with a small group of year 7 and 8 students working at lower than expected academic levels. The current study is based on Cognitive Load Theory and research suggesting that counting strategies overload working memory. A mixed-methods approach was used to identify whether structured manipulatives improved the additive reasoning and, addition and subtraction fluency in a sample of ten participants. Participants attended after-school intervention sessions of 45 minutes for seven weeks. The intervention focused on teaching additive reasoning and fluency using structured manipulatives. Inferential statistical analysis showed a statistically significant mean improvement in participants' ability to answer simple addition and subtraction questions. Tests constructed to operationalise additive reasoning also showed statistically significant mean improvement. Participants answered diagnostic questions operationalising various aspects of additive reasoning. Individual differences in understanding of additive reasoning were observed, and the inverse relationship between addition and subtraction proved to be a challenging concept. Semi-structured interviews provided themes of valuing the intervention and the manipulatives used. Due to the size and design of this study, it is not possible to extrapolate findings to other learners. However, the study may provide directions for future research. Structured manipulatives may have a role to play in enabling learners to begin to learn additive relationships and further securing recall of addition and subtraction facts. Students at years 7 and 8 may still need considerable exposure to additive concepts; moreover, returning to manipulatives may develop this knowledge. Finally, the findings from the diagnostic questions help show the complexity of additive reasoning. Classroom practitioners may need to further develop their knowledge of additive reasoning, its importance, and the individual differences and misconceptions that learners hold in order to provide considered learning experiences.

Acknowledgements

I would like to thank Robin Averill for her support completing this thesis. Her supportive and knowledgeable feedback was invaluable.

Joanna, Thomas and Emma – thank you for your patience. This is for you.

Table of Contents

List of Tables

List of Figures

Abbreviations

CLT	Cognitive Load Theory
COF	Conceptually ordered fluency
DCM	Digits Correct Per Minute
LTM	Long-term memory
MCD	Mean Correct Digits
NDP	Numeracy Development Project

Chapter 1

Progress in mathematics learning is of concern to learners, parents, and teachers, as well as to governments. Fundamental to progress in mathematics learning is a strong foundation in numerical strategies, skills, and understanding. This thesis will examine two areas of mathematics (additive reasoning and arithmetic fluency) through the lens of Cognitive Load Theory (CLT). The current study involves a short intervention aimed at improving additive reasoning and arithmetic fluency using structured manipulatives.

1.1 Overview of Chapter 1

Chapter 1 will outline the importance of achievement in mathematics both to individuals and economies. New Zealand's mathematical performance, in comparison to other countries and relative to its own standards, will be shared. Specific areas within mathematics have been shown to predict later mathematical achievement and these will be discussed. The varied data examining manipulative use as a pedagogical instructional method will be discussed and gaps in the research identified. The current study is based upon research into the capacity of working memory, and CLT. Research around these areas will be shared. Finally, the researcher will outline his previous experiences and will discuss how these have shaped the study design and aims. The research questions for the current study will be shared.

1.2 The importance of mathematical achievement

Research has shown links between educational achievement; income in later life; and life expectancy (Hahn & Chattopadhyay, 2019). Poor numerical skills have been linked with a range of negative social outcomes such as poor mental health, and a lack of financial security (Bynner & Parsons, 2006). Further studies have found links between mathematical achievement and economic prosperity of a nation (Baumann & Winzar, 2016). Therefore, what contributes to mathematical achievement, and the variance in countries' mathematical achievement, are important areas for research.

1.3 Mathematical achievement – the New Zealand context

New Zealand students' mathematical performance can be described as middling in comparison to that of their international peers (Ministry of Education New Zealand, 2017). However, further examination shows that at years 5 and 9, 41% and 42% respectively of New Zealand students are working below the Trends in International Mathematics and Science Study benchmark for these ages (categorised as *below low* or *low*). The proportion of students with achievement categorised as *below low* or *low* is less than the international median at year 5 but not at year 9. This is largely due to the greater proportion of lower achieving countries participating in the year 9 assessments (Ministry of Education, 2017). Furthermore, statistics from internal New Zealand assessments suggest that one third of students are not at the expected national curriculum level by the end of year 8. Students do not appear to catch up as the mean progress students make decreases with each year of education (Ministry of Education, 2018).

In summary, data from international and national assessments suggest that New Zealand has a persistent tail of underachievement (Ministry of Education New Zealand, 2017). Identifying what characterises, and how to address low mathematical achievement is important for individual learners and educational policy locally, nationally, and internationally.

1.4 The role of arithmetic knowledge and mathematical reasoning in mathematical achievement

There is a growing body of research focusing on understanding key contributors to mathematical achievement (Gilmore et al., 2018). Arithmetic knowledge is defined as the ability to calculate, while mathematical reasoning is defined as knowledge of the relationships between quantities (Nunes, Bryant, Barros, & Sylva, 2012). A longitudinal study completed by Nunes et al. (2012) researched these two types of mathematical knowledge (arithmetic knowledge and mathematical reasoning). The study, completed in the United Kingdom, lasted five years and involved over 1600 eight to 14 year-olds. Both arithmetic and mathematical reasoning were found to be independent predictors of later mathematical achievement. However, scores in mathematical reasoning tests devised by the researchers predicted later mathematical achievement in national tests more closely than arithmetic scores (Nunes et al., 2012). Parallels can be found in analysis of New Zealand underachievement. Chamberlain (2013), when investigating the weak areas of a number of New Zealand students, states that adding and subtracting whole numbers is a weakness for students in the low achievement band at year 5, while at year 9 students had difficulty applying number facts and procedures (Section 2.7).

It is important to find effective teaching methods and tools to lift achievement in mathematics; the areas of arithmetic knowledge and mathematical reasoning appear to be worthy of consideration. The current study will examine one aspect of mathematical reasoning termed additive reasoning, alongside fluency with addition and subtraction facts.

1.5 Manipulatives as pedagogical tools

A body of research examines what constitutes best practice in the mathematics classroom, with the use of manipulatives identified as being important for the development of mathematical understanding for many students. Bouck and Park (2018) define manipulatives as "concrete objects that can be used to help students understand and solve mathematics problems" (p. 66). Manipulatives are often identified as an integral part of educational policy for mathematics both in New Zealand (Anthony, 2007) and abroad (Department for Education, 2013; Ministry of Education, 2012). Research examining the use of manipulatives has produced a range of findings. Sowell (1989) completed a meta-analysis of 60 studies examining manipulative use, and identified a moderate-to-strong effect on learning when manipulatives were used over a significant period of time (one year). However, a more recent meta-analysis of 55 studies with a sample of 7,327 learners identified only small to moderate effects of manipulative use moderated by other factors such as manipulative design, instruction, and mathematical topic (Carbonneau, Marley, & Selig, 2013). Sowell (1989) called for further research examining the impact of specific types of manipulative in particular situations. Although the term situations might refer to a range of variables, the current study will examine the effectiveness of a specific manipulative type on learning in relation to additive reasoning and fluency.

The current study examines the use of structured manipulatives. For the purpose of this study, the term "structured manipulatives" will be defined as, "manipulatives that accurately illuminate the quantities and relationships in additive and multiplicative relationships." A search of the literature was unable to find the term "structured manipulatives" or any synonymous terms; therefore, the definition was developed as part of the current study. The current study aims to add to the body of research on manipulative types. The two structured manipulatives used are Numicon shapes and Cuisenaire rods (Figure 1.1).

In summary, research examining manipulative use has reported inconsistent findings. The current study aims to explore the use of two structure manipulatives on additive reasoning and fluency.

Figure 1.1. Cuisenaire rods (left) and Numicon shapes (right)

1.6 The role of working memory and Cognitive Load Theory on learning

The theoretical model that the current study draws from is CLT (Chandler & Sweller, 1991). CLT is an instructional theory centred around research of cognition and memory. Before describing CLT, a model of short-term memory which is foundational to CLT will be described. Baddeley and Hitch (1974) proposed a model of short-term memory that attempts to identify the key components within short-term memory; this is termed the working memory model. Working memory has a limited capacity (Miller, 1956) and contains three components: the phonological loop, visuo-spatial sketchpad, and the central executive. The phonological loop deals with auditory information, the visuo-sketch pad operates with visual information, while the central executive is believed to operate as a control centre mediating between the slave systems (Baddeley, 2018).

Specifics of CLT will be outlined before linking this theory to the present study (Section 1.7). CLT proposes that consideration of this cognitive architecture is vital when exploring the processes of learning. The relative difficulty of learning a new concept is moderated by the intrinsic load of the concept (the complexity of the concept) and the extraneous load (presentation of information) (Sweller, Ayres, & Kalyuga, 2011). A further assertion of CLT is that long-term memory (LTM) has an important role to play in skill acquisition and learning; once key basic skills are embedded in LTM, these can be retrieved without placing undue load on working memory (Sweller & Chandler, 1994); schemas defined as, “cognitive constructs for organising information” (p. 186) are stored in long-term memory. Sweller and Chandler (1994) state that schemas allow us to ignore the variability and specifics of a concept and increase the ability to generalise. This reduces the load on working memory.

Sweller and Chandler (1994) make a distinction between material that is difficult to learn and material that creates a heavy cognitive load. Material that is difficult to learn may have many individual elements. For example, learning vocabulary from another language has many individual elements and is therefore difficult to learn. Material that places a high cognitive load has elements which interact, such as learning the grammatical rules of another language (Sweller & Chandler, 1994). Therefore, learning the grammatical structures of another language has a high intrinsic cognitive load.

The instructional choices of a teacher may place increased cognitive load on a learner, termed extraneous load. Consideration of instructional choice also allows teachers to find ways to decrease working memory load. One example, the modality effect, suggests that presenting information in different modalities (linked to the components of working memory) will decrease cognitive load. For example, information presented visually and verbally can decrease cognitive load (Sweller et al., 2011).

In summary, CLT uses cognitive psychology research findings to explain the challenge of learning in terms of difficulty and cognitive load. It also allows consideration of instructional choices that may increase or decrease cognitive load.

1.7 How Cognitive Load Theory informs the current study

The current study draws from cognitive load theory in several ways. First, it is conjectured that the low-achieving participants selected for this study have not mastered basic addition and subtraction facts. These learners may use less sophisticated methods for calculation, such as counting procedures. During mathematics lessons this may lead to an increased cognitive load as learners are challenged with a wide range of elements that may interact such as mathematical vocabulary, structure of a problem, number of steps to solve a problem, and calculating the equation.

Secondly, it is possible these participants do not have a part-whole schema for additive reasoning. Therefore, they are likely to be unable to see the connection between related addition and subtraction facts. For example, a learner with sophisticated part-whole schema and memorised facts might be able to recall 7 – 4 from LTM using knowledge of 4 + 3 = 7 and an understanding of the relationship between addition and subtraction. Without mastery of these skills and concepts, learners need to find a less sophisticated method for subtraction, such as a counting strategy (Section 2.9). Once again, this can increase cognitive load in a classroom setting.

In summary, lack of additive reasoning, and mastery of addition and subtraction facts, may result in an increased cognitive load. This regression to a counting method is particularly relevant for the New Zealand context.

1.8 The importance of counting as a foundation of arithmetic knowledge in New Zealand

In 2000, Numeracy Development Projects (NDP) were introduced with the intention of raising the standard of mathematics teaching (Higgins & Parsons, 2009). As part of the NDP, a Number Framework provided teachers with a detailed and structured learning progression, tasks, pedagogies, and assessments to progress students through sequential areas of number (Ministry of Education, 2008). The first four stages of the Number Framework focus on counting (Ministry of Education, 2008). Young-Loveridge (2011b) suggests that the projects had a narrow focus on counting strategies, possibly at the expense of learners' focus on subitising and unitising. Some students may have become stuck at the counting stages due to regular explicit teaching of counting strategies rather than being progressed to part-whole understanding (Young-Loveridge, 2011b). Similar considerations have been made in other countries; Nunes, Bryant, Hallett, Bell, and Evans (2009) state that additive reasoning is not given adequate weight in the United Kingdom curriculum.

1.9 Using structured manipulatives to develop arithmetic knowledge and fluency

The modality effect found in relation to CLT (Section 1.7) states that teaching using a combination of visual and verbal information may decrease cognitive load. This idea supports the use of manipulatives in the current study. It is hoped that structured manipulatives may provide a visual tool to enable participants to learn basic addition and subtraction facts and develop understanding of additive reasoning (part-whole schema).

Manipulatives emphasising, or limited to, one-to-one correspondence such as counters or beans may have a different effect on learning compared to manipulatives that are structured (Section 1.5). This is a gap in the literature. It is an assumption of the current study that unstructured manipulatives may promote the use of counting strategies. Using unstructured manipulatives, the parts are always equal to one. The whole can be found using count all or count on strategies. It seems plausible that manipulatives in groups of one may not accurately represent the properties of additive reasoning. Providing learners with manipulatives that emphasise part-whole relationships may promote additive reasoning understanding.

1.10 Researcher background

I recognise that my experiences create a worldview that informs my research. My professional background provides a broad context for my study. I taught in the United Kingdom for ten years, leading mathematics teaching and learning at a primary school for four years. In 2014, the United Kingdom introduced a curriculum with a greater focus on teaching for conceptual understanding (Department for Education, 2013). Part of my role was to oversee the implementation of this curriculum in my school and the increased pedagogical focus on teaching conceptual understanding, particularly through manipulative use. A few years later, I moved to New Zealand and in January 2018 I started teaching a composite year 7/8 class. While teaching this class, I became aware that a group of students lacked conceptual understanding across a range of mathematical concepts. I also noticed that this group of students used their fingers to count. Subtraction was a particular challenge, and they were unable to see the relationship between addition and subtraction. I felt that these students would benefit from experience with manipulatives and the opportunity to learn some of the fundamental concepts of mathematics. I wanted to design and implement an intervention to study how I could assist the group of students exhibiting these gaps in mathematical understanding that could possibly inform practice across the school for students with similar difficulties.

The approach to this study was also informed by my previous academic experiences. My first degree was in psychology, focussing heavily on cognitive science. I have taught primary school aged children for 13 years and I have always been interested in effective teaching techniques. Although I understand that learners bring prior learning, misconceptions, cultural, social, attitudinal and, genetic influences to learning, I believe that there are teaching methods and practices that are more effective (or more likely to be effective) than others. This belief is grounded in my academic experiences of cognitive psychology (Baddeley, 2018; Miller, 1956; Sweller et al., 2011). Throughout most of my career, I have attempted to seek a greater understanding of what those practices might be.

1.11 Summary

In summary, given the importance to learners of sound mathematical understanding and recall, the national and international research showing the importance of addition and subtraction facts, and additive reasoning, my personal interests and experience, exploring structured manipulative use for advancing additive reasoning and fluency became the key driver of this thesis research. The current study hypothesises that a reliance on counting strategies and a lack of understanding of additive reasoning in low-achieving year 7 and 8 students may result in an overloaded working memory. These students may have a reduced ability to identify relationships between quantities. Structured manipulatives may help participants to develop understanding of additive reasoning, and move participants away from counting strategies.

Specifically, this study sets out to explore the following research questions:

1) Does the use of structured manipulatives improve fluency of addition and subtraction facts?
2) Does the use of structured manipulatives improve additive reasoning in this sample?
3) What perception of using structured manipulatives do year 7 and 8 students have?
4) Is there a link between learning gains and perception of manipulatives?
5) Are counting strategies frequently used by participants in this sample?

Chapter 1 has indicated that additive reasoning and addition and subtraction fluency are important for broader mathematical achievement. CLT has been discussed and linked to additive reasoning and fluency. The use of manipulatives as an instructional tool has been linked to the modality effect aspect of CLT. The rationale behind the choice of manipulatives has been shared. Finally, the researcher has outlined his background and has identified how this helped shaped the research. This context frames the current study in which structured manipulatives are used as part of an intervention to develop the additive reasoning, and addition and subtraction fluency, of currently low achieving year 7 and 8 students.

Chapter 2 will highlight and critically analyse the body of literature on addition reasoning, fluency, and manipulative use. Chapter 3 will discuss the research paradigm and worldview used for the current study. The choice of research method and data collection tools will be shared along with a rationale for these choices. Chapter 3 will also outline how the intervention was completed and how the collected data were analysed and coded. Chapter 4 shares the results of the current study, including descriptive and inferential statistics. Chapter 5 discusses the results of the study, links these to the research literature, and discusses implications of the study findings.

Chapter 2

The current study focuses on the development of addition and subtraction fact fluency, and additive reasoning. The literature review examines the research in both areas, beginning with additive reasoning. This chapter will also examine research identifying the relationship between additive reasoning and fluency. Furthermore, the literature examining manipulative use and CLT will be shared.

2.1 Defining, and examining the role of additive reasoning.

Nunes et al.'s (2012) work evidences the importance of mathematical reasoning as a predictor of mathematical achievement. Mathematical reasoning can involve understanding of additive, multiplicative and spatial relationships (Nunes et al., 2012). This literature review focuses on one aspect of mathematical reasoning - additive reasoning. The teacher/researcher completing this study identified additive reasoning as a focus for the struggling students in his class, some of whom were invited to participate in the current study. This focus on additive reasoning is supported by analysis of New Zealand students' performance in international tests (Section 1.3).

A range of terms appear to be synonomous with additive reasoning: part-whole relations (Resnick, 1992), arithmetic principle knowledge (Prather & Alibali, 2009); and understanding of addition and subtraction concepts (Canobi, Reeve, & Pattison, 1998). The current study will use the term additive reasoning throughout. The definition of additive reasoning used for this study is, "based on quantities connected by part-whole relations. Two central properties of part-whole relations involve (a) commutativity and (b) the inverse relation between addition and subtraction" (Ching & Nunes, 2017b, p. 483). Furthermore, as a function of additive reasoning, learners should understand that four expressions can describe the relationship between three quantities ($a + b = c$, $b + a = c$, $c – a + b$, and $c – b = a$), and three related expressions can be derived from a fourth given expression (Ching & Nunes, 2017b).

Ching and Nunes (2017b) suggest that understanding of additive reasoning is a crucial component of mathematical understanding. A recent study demonstrated the importance of additive reasoning on mathematical achievement; non-Causasian seven-year old students had an understanding of additive reasoning that was strongly related to mathematical ability, independent of working memory, and counting ability (Ching & Nunes, 2017b). Knowledge of additive reasoning may also have benefits on areas of mathematics more widely. Baroody, Ginsburg, and Waxman (1983) highlight how knowledge of mathematical structure can reduce the effort required when carrying out calculation. It might be argued that the effort involved is akin to working memory overload (Section 1.6). Another crucial affordance of mastering additive reasoning is the illumination of the structure behind real-world addition and subtraction contexts, allowing students to access, and be users of the mathematics of their world (Carpenter & Moser, 1984).

2.2 The development of additive reasoning

Research indicates additive reasoning develops progressively (Ding & Auxter, 2017). This assertion is supported by Canobi (2005) who found individual differences in young children's conceptual profiles of additive reasoning. Resnick (1992) discusses the importance of reasoning with mathematical entities, describing a four-layer theory of increasingly abstract mathematical knowledge progressing from reasoning with physical quantities (protoquantities) where comparisions of size or quantity are made. The second layer involves reasoning with and about quantities. At layer three, a shift in the use of numbers from adjectives to nouns occurs, and numbers are used abstractly. The final layer of formalised thinking is the *mathematics of operators* where operations can be reasoned with. This progression is consistent with the work of CLT, identifying how schema may become increasingly more sophisticated. Resnick (1992) specifically relates these layers to additive reasoning. At the first protoquantitive layer, learners understand that, "a whole quantity can be cut up into two or more parts and that the parts can be recombined to make the whole" (p. 49). In the second and third layers, these relationships persist and learners begin to enumerate quantities of objects. In the final layer, the learner is able to reason more abstractly with part-whole relationships. Resnick (1992) highlights that the relationships between the parts and the whole remain the same throughout the four layers. Although Resnick's work provides a pathway for developing reasoning and descriptors for complexity of thinking, it does not discuss how students can be taught these challenging but vital concepts or have moved from layers in a specific domain. Resnick's model provides further weight for the use of structured manipulatives for developing part-whole thinking, as each physical manipulative can represent a part and can be combined with another to create a whole (Section 1.9).

Research into additive reasoning shows variation in how easily sub-concepts are acquired. The commutative law of addition appears to be acquired relatively easily by learners (Canobi, Reeve, & Pattison, 2002). Baroody et al. (1983) suggest that children may discover commutativity informally and through play or early counting experiences. The inverse relationship between addition and subtraction appears to be a more challenging concept (Baroody, 1999; Baroody et al., 1983; Canobi, 2005; Koponen et al., 2018; Nunes et al., 2009). Torbeyns, Peters, De Smedt, Ghesquière, and Verschaffel (2016) studied understanding of the complement principle with a sample of 67 nine to ten year-olds. Participants were provided with a task where previous addition problems could be used to solve target subtraction problems. It might be expected that participants could use the visible addition problem and knowledge of additive reasoning to solve the target subtraction problem. Despite participants being specifically told that looking back may help them solve problems, only 12% of participants did so consistently. This demonstrates the challenge of the inverse relationship between addition and subtraction for some learners. Canobi (2005) refers to Piaget and suggests this may be due to learners identifying the positive characteristics more easily than the negative aspects.

In summary, the literature identifies the importance of additive reasoning as a facet of mathematical understanding and that some areas of this concept are learnt more easily than others. The literature offers parallels to the importance of schema discussed as part of CLT (Section 1.7).

2.3 Research on manipulative use for teaching additive reasoning

Several studies have examined how additive reasoning might be taught effectively; however, the findings have been inconsistent. Baroody (1999) found that an intensive training intervention was not effective in teaching first graders the complement principle. Nunes et al. (2009), however, completed an intervention which improved participants' understanding of the complement principle compared to the performance of a control group. In their study, visual demonstration (using unifix blocks) of the complement principle yielded a greater improvement compared to oral instruction. However, it may be difficult to distinguish between the impact of the unifix blocks and the modelling provided. A more recent study with kindergarten children showed that an intervention with unifix blocks led to significant greater understanding of the inverse relationship between addition and subtraction than a control group, and a group taught without manipulatives, suggesting the importance of manipulatives as part of an intervention (Ching & Wu, 2019). Ding and Auxter (2017) state that young children should be taught about additive reasoning using concrete materials, concluding this after examining the strategies that young children use when solving additive problems. However, in their study, the test instruments were not piloted and were distributed and administered by 35 teachers enrolled on a postgraduate course, so it is possible that not all tests were administered the same way.

In a theoretical model, Resnick (1984) shares a visual part-whole coding system where the two parts equal the length of one whole. This visual model closely replicates some of the manipulatives used in the current study (Cuisenaire rods), and hence is particularly relevant to the current study, supporting the rationale to use structured manipulatives to develop learners' understanding of additive reasoning. Comparison of Figure 1.1 and Figure 2.1 shows the similar affordances of the two manipulatives used in the present study and Resnick's model.

Figure 2.1. Modified from Resnick's (1984) visual part-whole coding system (p. 116)

However, despite some researchers highlighting the valuable role of manipulatives for teaching additive reasoning, this view is not unanimous. When discussing the difficulty of using direct instruction to teach the complement principle, Baroody (1999) suggests a range of strategies that may aid the discovery of this principle but does not mention the use of manipulatives. Baroody (1999) does suggest explictly using the terms "part" and "whole" when referencing quantities in an equation, a technique used in the current study as a result.

In summary, research on instruction of additive reasoning is inconsistent. Manipulatives may have a role to play in the development of additive reasoning, particularly if the manipulatives emphasise part-whole relations. Further detail on manipulative use will be shared in section 2.14.

2.4 The development of arithmetic knowledge

Baroody and Dowker (2003) identify some of the early 20th century theorists' views on the teaching and development of arithmetic; specifically, the skills versus meaning focus underpinning instructional programmes. Skills-focused theorists value drill, procedural learning of algorithms and number facts; however, meaning-focused theories, such as Brownell (1945), espouse an instructional view with conceptual understanding as a key component. Baroody and Dowker (2003) discuss four possible views on instructional programmes: skills first, concepts first, iterative development, and simultaneous development. A skills first approach values the transmission of key skills and facts with little or no thought for understanding. Conversely, a concepts first approach values meaningful memorisation of facts. Sitting between these two polarised views are iterative development and simultaneous development. Iterative development believes that one form of knowledge (conceptual or procedural) might lead to advances in the other (e.g., a counting procedure might allow learners to discover that addends are order-irrelevant). Finally, simultaneous development suggests that the two knowledge types build concurrently. Baroody's summary of these viewpoints informs the current study in two ways. Firstly, the current study aims to develop additive reasoning and fluency in a sample of learners and therefore the instruction used will be placed somewhere on the procedural to conceptual continuum. Secondly, it might be argued that teaching additive reasoning may itself by a form of conceptual knowledge (Section 2.12) and that development of this conceptual knowledge leads to improvements in procedural understanding of number facts. This review will now focus on the specific development of arithmetic and examine the literature surrounding this before looking more closely at the relationship between additive reasoning and fluency.

2.5 The role of counting in developing arithmetic knowledge

Arithmetic knowledge has its roots in early mathematical experiences with counting, and early arithmetic procedures may be the adaptation of counting skills (Geary, 1994).

Geary (1994) summarises five basic addition strategies that develop in complexity, and highlights the cross-cultural consistencies of these strategies: using manipulatives, finger counting, verbal counting, deriving facts, and fact retrieval. Carpenter, Hiebert, and Moser (1983) completed a three-year longitudinal study with 144 participants examining strategy choice for word problems with a variety of structures. These researchers found counting strategies were prevalent at younger ages; however, counting strategies were superseded by retrieval strategies for the majority of students by grade three. Strategy use was dependent on the structure of the word problem.

Geary (1994) describes how counting strategies vary in sophistication from a count-all strategy to a count-on from the first addend, then finally count-on from the larger addend. The increase in sophistication links to an increase in cognitive demand: count-on strategies involve a double-count. For example, solving 6 + 4 using a count-on from the first addend strategy would require counting the complete total and the total of the four to be added, *7 (1), 8 (2), 9 (3), 10 (4).* Research has shown that some students find this demand more challenging than others (Section 2.7). Geary (1994) states that once learners are able to use verbal counting strategies to solve addition facts, the next learning stage involves using known facts and deriving facts with small addends. Derived facts may begin from knowledge of doubles; as research has shown these are memorised more easily (Ashcraft, 1992; Carpenter & Moser, 1984).

2.6 The role of conceptual understanding in arithmetic development

Baroody (1994) describes two varying theories of arithmetic (fact-retrieval and schema-based). It is generally accepted that learners move from counting procedures to retrieval; however, it is debated whether fact retrieval is due to associations (Siegler, 1987) or driven by conceptual understanding (Baroody, 1999).

Gray and Tall's (1994) work supports the role of conceptual understanding in arithmetic, introducing the idea of a procept – the duality of a process and a concept. Specifically, 3 + 2 might be considered a process but could also be a concept itself; hence, the portmanteau procept. The idea of a procept is described in relation to arithmetic development. When using a count-all strategy, the learner is carrying out a process. The shift to counting-on strategies signifies that the starting addend is a procept (or whole) and the second addend is counted-on which could be considered a process. Once both addends are considered procepts, this becomes a known fact. Gray and Tall (1994) assert that this type of situation illustrates not a known fact learnt by rote, but a flexible form of understanding from which further facts can be derived. Gray and Tall (1994) emphasise that this shift in arithmetic thinking is a shift from procedural methods to a more flexible conceptual understanding. This idea is supported by Hopkins and Egeberg (2009) who discuss how using known facts to derive facts is underpinned by flexible use of existing knowledge; for example 5 + 7 might be calculated by decomposing 7 into 5 and 2. 5 + 7 then becomes 5 + 5 + 2.

In summary, although there is some debate over the role of conceptual knowledge in arithmetic development, a large body of research supports the idea of a conceptually based, flexible understanding and teaching of arithmetic.

2.7 Difficulties with counting and the impact on arithmetic knowledge

Despite the literature identifying some clear pathways for developing basic fact knowledge, a proportion of learners do not acquire this fluency and this impacts on their achievement (Chamberlain, 2013). Therefore, it is salient to examine the literature on learners who struggle in this area. Young children with poor mastery of addition and subtraction facts persist with finger counting when compared to students with mastery of these facts (Jordan, Hanich, & Kaplan, 2003). Counting strategies may frequently be used in the early stages of a learner's arithmetic knowledge, or may persist in low-achieving learners (Gray & Tall, 1994; Siegler, 1987). Dowker (2005) theorises that an overreliance on counting strategies may impede arithmetic development. This reliance may inhibit the development of more sophisticated strategies due to the attentional requirements of counting; furthermore, Dowker (2005) suggests that learners who rely on counting procedures to calculate also find counting cognitively taxing.

Gray (1991) examined the strategy choices of 72 learners who had been categorised by their teachers as below average ability, average ability, or above average ability. Above average students recalled facts more frequently and were able to derive facts from known facts more frequently when compared to below average peers. Gray suggests that less sophisticated strategies chosen by below average learners provide security and discourage learners from storing the fact in LTM. Supporting this, Ostad (1997) compared students with mathematical difficulties to their peers, finding that students with mathematical difficulties are more likely to use less sophisticated strategies.

2.8 The relationship between working memory, counting, and arithmetic

The present study is framed around CLT, suggesting that consideration of cognitive load is important to learning. Research has found that working memory, alongside intelligence, predicts academic achievement (Schneider & Niklas, 2017). Linked specifically to mathematics, research has found that difficulties with arithmetic are linked to deficits in working memory (Adams & Hitch, 1997; Jesica, Silvia, Irene, & Juan Pablo, 2018; McLean & Hitch, 1999). Research has also reported an interaction between working memory, anxiety, and mathematical performance (Ganley & Vasilyeva, 2014). Further evidence from other domains shows that counting tasks can be used successfully to test working memory (Van Den Hout et al., 2010). de Chambrier, Thevenot, Barrouillet, and Zesiger (2018) found that children who used finger counting more frequently had lower working memory capacity. It could be hypothesised that finger counting lowers demands on working memory, and supports the learner in the short-term; however, finger use may promote the perpetual use of counting strategies in some learners. Gray (1991) states that learners who perpetually count do not have the feedback loop that allows a network of known facts to be learnt, and subsequently learners are unable to use new facts to derive further facts. The ability to derive facts then leads to a greater number of known facts. Although counting is often considered the gateway to calculation, there may be other pathways (Findell, Kilpatrick, & Swafford, 2001; Young-Loveridge, 2011a).

2.9 Alternative pathways to arithmetic understanding

The NDP placed part/whole thinking after counting strategies in the broad progression of number knowledge; however, researchers have called for a greater emphasis on using part/whole thinking to develop arithmetic knowledge (Young-Loveridge & Bicknell, 2015). It is important to consider the focus on taught counting strategies in the NDP alongside the finding that some learners become stuck in using counting to calculate (Section 2.8). One possible pathway to fluency that requires greater research is the use of subitising (Young-Loveridge, 2002). Subitising can be defined as instantly seeing an amount without counting (Faulkner & Ainslie, 2017). Clements, Sarama, and MacDonald (2019) refer to subitising as the neglected quantifier and suggest that conceptual subitising (subitising two parts and creating a whole) may have an organising role in part-whole understanding alongside that of counting. Specifically, learners might subitise two and then three, then combine these two parts to identify the whole. A developmental progression would be to subitise five because of the identification of the sub-groups two and three. The importance of understanding groups and structure was identified in a recent study by Kullberg and Björklund (2019), who noted that young children could solve 3 + ? = 8 by using fingers to identify part-whole relations rather than counting; finding that identifying parts, wholes and structure, can have a powerful impact on arithmetic knowledge. The role of subitising and structure is relevant to the current study as Numicon shapes are organised into patterns which may encourage similar processes to conceptual subitising. Relevant to the New Zealand context, a concluding remark from Young-Loveridge (2011b) stated that subitising should have been included alongside counting in New Zealand's NDP.

In summary, although most of the literature suggests counting is the foundation of arithmetic, it is possible that other pathways exist and research into effective teaching methods relating to these pathways is valuable.

2.10 Counting and working memory – a CLT perspective

CLT (Sweller & Chandler, 1994) supports the notion that classroom practice should focus on embedding basic skills (such as arithmetic knowledge) in LTM to free working memory. In a study of 228 primary school children, Geary, Hoard, Byrd-Craven, and Catherine Desoto (2004) found that working memory importance decreases with age as basic facts become stored in LTM; an additional finding was that learners with mathematical difficulties also have difficulties with working memory. Cognitive psychology also offers insight into this area using various data collection methods. Cho, Ryali, Geary, and Menon (2011) used brain imaging techniques with 7 to 9 year-old students and showed different parts of the brain were activated when students used retrieval strategies compared to counting, suggesting that counting strategies use different cognitive resources to retrieval strategies. Furthermore, a review of brain imaging research on arithmetic learning highlights that automatic retrieval lessens working memory load (Zamarian, Ischebeck, & Delazer, 2009). Geary et al. (2004) suggest that deficits in working memory lead to inaccurate and slow counting procedures and, as a result of this, basic facts are moved into LTM more slowly. This assertion is slightly at odds with Gray (1991) who suggests that counting procedures may provide learners with security. Another consideration for the perseverance with counting strategies is teacher validation; specifically, the focus on counting procedures in the New Zealand Number Framework may have resulted in some practitioners overemphasising counting methods to calculate (Young-Loveridge, 2011b). Cheng (2012) states that children who are encouraged to count can become reluctant to move to more sophisticated strategies. Despite the lack of consensus regarding how counting can limit the development of retrieval strategies, a pragmatic view would suggest that finding effective ways to develop learners to use more sophisticated strategies is of importance.

In summary, it is important to find methods to increase retrieval of basic facts in order to reduce cognitive load. The current study aims to examine if structured manipulatives might shift participants from counting strategies to retrieval strategies, possibly using similar processes to subitising. Alongside this, a body of literature suggests developing conceptual knowledge (additive reasoning) alongside arithmetic is desirable for creating flexible understanding of arithmetic.

2.11 The distinction between fluency and arithmetic

Arithmetic knowledge was previously defined as the ability to calculate (Nunes et al., 2012). This definition might be considered ambiguous and challenging to quantify as calculation can use a range of strategies of varying sophistication, ranging from a count-all strategy to retrieval of a fact. The current study uses the term fluency to refer to the accuracy and speed of calculation with addition and subtraction facts. Wubbena (2013) links fluency to mastery understanding by combining accuracy and speed of calculation and identifies that measuring the construct arithmetic without speed can overestimate learners' abilities. The rationale to include speed to measure fluency rather than arithmetic supports the model of CLT where the importance of automatic recall of facts from LTM has been presented (Section 2.10).

2.12 The interaction between fluency and additive reasoning (procedural and conceptual understanding)

A large scale research study (Nunes et al. 2012) has shown that mathematical reasoning and arithmetic ability (fluency) make independent contributions to later mathematical achievement; furthermore, these two constructs were significantly moderately correlated. Nunes et al. (2012) make the assertion that reasoning and arithmetic should be considered different constructs and state that mathematical reasoning is a form of conceptual knowledge. Conceptual understanding is defined as, "the implicit or explicit understanding of principles that govern a domain and of the interrelations between units of knowledge within a domain"

(Rittle-Johnson, Siegler, & Wagner Alibali, 2001, p. 346). Procedural knowledge is defined as, "the ability to execute action sequences" (p. 346). Using this definition, being fluent with addition and subtraction facts may be identified as a form of procedural knowledge. Research will be discussed which identifies links between counting strategies and lack of additive reasoning. Therefore, it is important to consider if and how these two constructs interact.

Additive reasoning and arithmetic are interwined and related concepts. For example, calculating 21 – 19 might be most efficiently solved using the inversion principle and counting on from 19. What is unclear is how one construct informs the other (Canobi, 2009; Ching & Nunes, 2017b). Baroody et al. (1983) found that participants who used counting strategies were less likely to identify the complement principle than participants who used more sophisticated strategies; however, this was not true of the commutative law which appeared to be available to all learners. Research on a smaller scale has found that ability to retrieve addition and subtraction facts was related to conceptual knowledge of additive reasoning (Canobi et al., 1998). In a study of 200 second and third grade students, Cowan et al. (2011) found that addition and subtraction fluency was highly correlated with measures of conceptual understanding; however, the measures of conceptual understanding included, but were not restricted to the commutative and complement principles.

Despite research showing the links between additive reasoning and fluency, the mechanisms for the interdevelopment of these two concepts are less clear. One possible explanation of the relationship is that understanding of additive reasoning helps learners to organise knowledge more efficiently, thus increasing fluency; alternatively, the ability to retrieve arithmetic facts might afford learners greater awareness of additive relationships (Canobi et al., 1998). Baroody (1999) suggests a model where learners with awareness of the principles of additive reasoning identify these principles before searching for the arithmetic triple that would provide the key to solving a calculation (Figure 2.2).

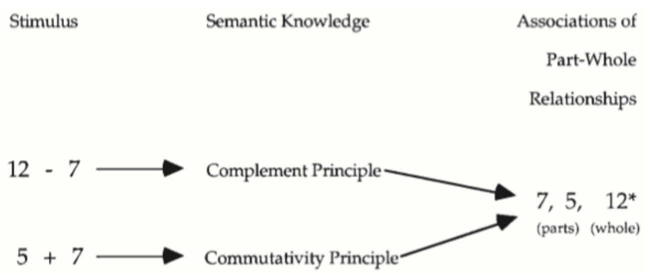

Figure 2.2. Baroody's (1999) model identifying how additive reasoning may aid retrieval (p. 152)

2.13 Studies designed to improve additive reasoning or fluency

Research has found inconsistent results relating to the improvement of fluency and additive reasoning. Although not specifically detailing additive reasoning, a twelve-week study with 69 participants with low levels of calculation skills showed that an intervention focusing on conceptual and procedural understanding significantly improved fluency compared to controls (Koponen et al., 2018). This provides evidence for the interdevelopment of procedural and conceptual understanding. Further support comes from Baroody et al. (1983) who devised a task where students who used additive reasoning principles achieved more highly in a game scenario. Alongside this, participants were coded on the efficiency of their addition retrieval. Participants used the commutativity principle regardless of the efficiency of their addition retrieval; however, the complement principle was only frequently used by participants who recalled the related addition fact from memory. "It is as if laborious calculation used up attention – leaving none for discovering and using complement principles" (Baroody, 1999, p. 167). These findings are consistent with CLT and the work of Gray (1991) where working memory is devoted to procedural methods, therefore resources are not available to identify additional structure.

Not all interventions successfully improve additive reasoning or fluency. Baroody (1999) found that a sample of gifted first grade students were unable to identify the relationship between addition and subtraction, and that an intervention was unable to alter this finding. One possible explanation given was that mastery of addition facts enables discovery of the complement principle; knowledge of addition facts might be required to reach a threshold level or to be practised extensively before the complement principle can be used (Baroody, 1999).

Development of additive reasoning may involve the understanding of rules. Baroody et al. (1983) point out that the use of shortcuts or rules may be viewed negatively by students (this may lead to the child persisting with an inefficient and mentally taxing procedural approach which perpetually blocks the ability to notice mathematical regularity). It is possible that over learning of addition facts may be required to enable students to identify additive reasoning principles. However, a recent study focusing on training specific addition facts did not lead to an increase in the ability to answer a complementary subtraction question, or improve the ability to answer untrained facts (Walker, Bajic, Mickes, Kwak, & Rickard, 2014). It should be highlighted that this study contained no specific or explicit teaching of additive reasoning, rather drill practice of addition facts. It would appear that procedural learning of addition and subtraction facts has limited benefits. This is supported by Gilmore et al. (2018) who found that conceptual understanding was highly correlated with mathematical achievement when the application of a range of additive reasoning principles was used to operationalise conceptual understanding in a group of typically developing preterm 8 to 10 year-olds.

Canobi (2009) completed a study involving practice of addition and subtraction facts. One treatment condition was composed of addition and subtraction questions that were presented in a specific order to illuminate the structure of additive reasoning, e.g. a + b followed by b + a. Participants in the conceptually ordered treatment made gains in retrieval of addition and subtraction facts and conceptual understanding which the random practice group did not make. The gains in retrieval were not evident in already practised problems, only unpractised problems. It should be noted that the gains in retrieval were not statistically significant. When placed alongside the work of Walker et al. (2014), this result adds to the literature regarding the links between conceptual and procedural understanding in the field of additive reasoning and fluency. Procedural practice of addition and subtraction facts appears to lead to very specific gains; conceptually ordering procedural work may have broader gains with other unpractised addition and subtraction facts. Mirroring Baroody's model (Figure 2.2), Canobi (2009) states, "a concept-based reorganization of children's memory of addition and subtraction problems and their answers could help to explain the finding that conceptual knowledge enhances children's use of retrieval-based procedures to solve unpracticed problems" (p. 147). It is important to note that these two studies (Canobi, 2009; Walker et al., 2014) operationalised procedural fluency in different ways. Canobi operationalised fluency by proportion of retrieval strategies while Walker et al. (2014) used proportion of correct questions within a timeframe.

In summary, the research suggests that teaching additive reasoning and fluency together may have benefits for each domain. Additive reasoning may be considered a form of conceptual understanding while fluency might be considered a form of procedural understanding. Two possible explanations for this mechanism have been discussed and these both support CLT. A more conceptually organised schema may allow for more efficient retrieval (Canobi, 2009) or, placing addition and subtraction facts into LTM memory creates working memory capacity to attend to additive reasoning relationships (Baroody, 1999) which subsequently allows access to a greater range of known or derived facts.

2.14 The role of structured manipulatives to develop additive reasoning

Section 2.3 briefly outlined the rationale for using structured manipulatives for teaching additive reasoning. Previous research has examined the inconsistent findings of manipulative use on learning. It is clear that a range of factors modulate any possible outcomes on learning (Carbonneau et al., 2013). Thompson (1994) discusses the importance of instructors considering the concept they wish to teach when selecting a manipulative. This is relevant to the current study as the manipulatives chosen may help to illuminate part-whole structures, and move participants away from counting strategies. Manches and O'Malley (2016) discovered that unifix blocks allowed participants to find a greater number of solutions in a partitioning task (attempting to find the partners for a given whole), and that manipulatives led to more sophisticated strategies being used. Bonne (2003) identified that some manipulative choices (the number line) encouraged learners to count in ones despite already having part-whole thinking. Bonne called for further research into the use of structured manipulatives to develop part-whole thinking. The use of manipulatives structured to replicate the base-10 system has been promoted (Bonne, 2003; Young-Loveridge, 1999). Mulligan and Mitchelmore (2009) emphasise the importance of pattern, structure and regularity in mathematical achievement, conjecturing that students with understanding of structure would be able to unitise, leading to improved number knowledge. The present study adds to the literature by using materials that allow each number to 10 to be unitised (Numicon and Cuisenaire rods).

Numicon equipment was developed in 1996 as part of a teaching programme aimed at making mathematics more concrete for learners (Wing & Tacon, 2007). Research on Numicon's efficacy is limited. A study with a sample of 25 students found no significant effect of the intervention on achievement; this study called for more research into the benefits of Numicon (Forder, 2016). The second manipulative used in this study, Cuisenaire rods, was invented in Belgium 85 years ago by George Cuisenaire (Kurumeh, 2010). Although the rods can be unitised to represent different amounts, initially each rod is used to represent one of the numbers 1 to 10. Each rod's value is equivalent to its length in centimetres; therefore, this structured manipulative has parallels with measurement. Some research has found an impact of Cuisenaire rods on the learning of specific mathematical concepts. In a study of 200 secondary school students, an intervention using Cuisenaire rods led to significantly higher post-test scores than students taught conventionally (Kurumeh, 2010). Green, Piel, and Flowers (2008) found that instruction with manipulatives (including Cuisenaire rods) significantly reversed preservice teachers' arithmetic misconceptions; however, this study did not use a control group. A review of the literature could not find any studies detailing the use of Numicon or Cuisenaire rods on fluency or additive reasoning with children. The search of the literature used the database SCOPUS and the terms Numicon or Cuisenaire with fluency, additive reasoning, part whole, and arithmetic, respectively.

2.15 Summary

This literature review has emphasised the importance of additive reasoning alongside fluency, suggesting that these two concepts are related. Proponents of meaningful, conceptual arithmetic instructional methods have been shared and it has been suggested that additive reasoning is a form of conceptual knowledge. The importance of counting has been discussed alongside the difficulties some learners have when counting to calculate. There is evidence that counting strategies may inhibit the development of more sophisticated strategy use of some learners, possibly due to the role of working memory in counting. Alternative pathways to arithmetic have been suggested, and the emphasis of counting in the NDP has been discussed. Inconsistent research into the development of additive reasoning and fluency has been shared. Finally, the possibility of structured manipulatives enabling learners to identify additive reasoning concepts and develop fluency has been shared. To highlight this, examining the manipulatives (Figure 1.1) alongside the models provided by Resnick (1984) (Figure 2.1) and Baroody (1999) (Figure 2.2) illustrates how these manipulatives might promote additive reasoning and fluency through the development of part-whole schema. The current study aims to explore whether an intervention using structured manipulatives to teach additive reasoning and fluency might lead to improvements in both areas. This is in alignment with proponents of meaningful, flexible, conceptually based models of arithmetic development.

Chapter 3

3.1 Reflecting on decision making in terms of theory and context

Creswell (2011) discusses the importance of researchers stating their philosophical worldview and assumptions when designing and carrying out a study. These assumptions should be made explicit, demonstrating what the researcher brings to the research (Creswell, 2011). In section 1.10 I discussed my belief that certain teaching methods may be inherently more effective than others. Moreover, I believe that quantifiable improvements in learning are outcomes from effective teaching methods. These beliefs, coupled with my background studying cognitive psychology, lead much of my worldview to being consistent with positivism and to strongly value the importance of statistical analysis in research (Newby, 2010).

Creswell (2011) refers to positivism as the scientific method with a focus on quantitative data collection. However, Newby (2010) also highlights the complexity of different philosophical worldviews and that these are not necessarily dichotomous. My experience of teaching leads me to also value a pragmatic approach. Given the teaching context of the study, the pragmatic worldview was the most dominant when designing this study and selecting mixed methods as a research method. Creswell (2011) states that "pragmatism opens the door to multiple methods, different worldviews, and different assumptions, as well as different forms of data collection and analysis" (p. 11). Furthermore, Patton (1990) promotes the use of mixed methods as an appropriate research method for the pragmatist worldview as it combines a positivist emphasis alongside a more grounded practical approach.

3.2 Overview and theory of mixed methods

Creswell (2011) discusses the value of mixed methods research in social science; it harnesses the advantages of quantitative and qualitative data collection methods and provides greater insight. Both quantitative and qualitative data collection methods have their advantages. Quantitative research has a focus on measurable and observable phenomena and uses inferential statistics to provide insight, and quantitative research is based on the premise that a study may be replicated, elaborating findings and allowing for the derivation of theory (Newby, 2010). Conversely, qualitative research methods can be used in natural settings and allow for relative truths dependent on situation and time: “The task of a qualitative researcher is not to look at how people behave as outsiders but to understand how individuals see the world.” (Newby, 2010, p. 119). Both data collection methods have weaknesses; Creswell (2011) states how mixed methods can neutralise these limitations to improve the quality of research. “By combining two (or more) research methods with different strengths and weaknesses in a research study, you can make it less likely that you miss something important” (Johnson & Christensen, 2014, p. 53). Not only is it important that a study design provides appropriate data for the research questions, it is also important that the method is suitable for the context. The present study has an educational context and Sammons (2010) asserts that mixed research in education is worthwhile as it offers the possibility of transfer to the classroom.

The mixing of methods when collecting data is becoming more prevalent; however, this type of design is not without controversy. Despite this, mixed methods use of different data collection methods (quantitative and qualitative) has been described as complementary (Johnson & Christensen, 2014), allowing exploration of an issue flexibly and in real time (Newby, 2010) which enables successful aspects of an intervention to be transferred (Drabble & O’Cathain, 2015). This is relevant for the current study as it is hoped that any findings from the research can be transferred to the classroom.

A defining characteristic of mixed methods research is the collection of both quantitative and qualitative data (Creswell, 2011). Moreover, the collection of the data using different methods is strategic (Section 3.3). Johnson and Christensen (2014) discuss that a fundamental principle of mixed methods is purposeful selection of data collection methods creating a study that is improved by the different strengths of the methods. Newby (2010) states that both data collection methods can be used to compare results and possibly reinforce findings.

Mixed methods was a suitable choice for the present study because both quantitative and qualitative data are required to answer the research questions. Specifically, quantitative data were collected to measure any change in participants’ fluency, and additive reasoning. Qualitative data were collected to examine participants’ perceptions of the study. Also, mixed methods adds rigour to the study as the study uses a small convenience sample. Generating conclusions from small convenience samples can be problematic so a mixed approach allows for triangulation (Creswell, 2011). This triangulation gives the researcher the possibility to corroborate findings or search for conflicting information (Johnson & Christensen, 2014). Finally, mixed methods is suitable for the size and context of this study;

Newby (2010) states that mixed methods are valuable when completing "action research, particularly where the research takes the form of an experiment involving pre-test, intervention and post-test and where the insights of more qualitative assessments can add considerable value" (p. 131).

3.3 Study design

In the present study mixed methods were used with a quasi-experimental design (Johnson & Christensen, 2014). A sequential explanatory design was selected because this enabled targeted data collection to aid exploration of research questions. This design suits researchers with a strong quantitative focus (Creswell, 2014) and meets the needs of the current study as research question 1 focuses on whether there will be an improvement in scores from pre-test to post-test. Within mixed methods research, there are a range of designs and orders that can be tailored to best fit a study. Sequential mixed methods designs have an ordered approach with one type of data collected first. Creswell (2011) discusses how this ordering is purposeful. Data collected in the quantitative phase provides lines of inquiry in the qualitative phase. Students were invited for the interview phase based on their performance in the quantitative phase.

After the intervention, the researcher conducted qualitative semi-structured interviews to explore students' lived experience of the intervention (Creswell, 2014). Participants were invited for interviews purposefully using maximum variation sampling. Four participants were invited based on mean fluency scores and variation in mean improvement in fluency scores. The rationale was to select participants with the largest variation of these measures and students who score more closely to the mean. Johnson and Christensen (2014) state that maximum variation sampling increases rigour by ensuring that all types of cases are selected, adding weight to themes that exist across the different cases.

3.4 Overview of data collection tools used

The quantitative phase of this study used an identical pre-test and post-test containing a range of test instruments (Appendices A, B, C, D), a decision informed by Johnson and Christensen's (2014) view that testing (using ratio scales) is a desirable data collection method in education for measuring performance. Johnson and Christensen (2014) discuss the strengths of standardised tests but also state that researchers might design test instruments to operationalise specific variables. The data collection tools in the current study were informed from previous research in this field, and then designed and refined through trialling by the researcher.

Piloting of data collection tools was completed in two stages. First, the researcher invited students who would not be part of the research to attempt the tests. These students were of average mathematical ability. Timings were checked to gauge how many questions were answered in the time provided, and perceptions of the test tools were obtained. Also, a teacher and group of students from a nearby school trialled the data collection tools. These students were the same age and had a similar mathematical ability to participants in the study, providing an opportunity to check wording and layout of the questions. The teacher offered some critical feedback on the data collection tools and how these were interpreted. This informed some revisions of the data collection tools.

A range of data collection tools was used in the present study (Table 3.1). Within the specific research field of additive reasoning, Prather and Alibali (2009) state, “The vast majority of studies utilize single-faceted knowledge assessments, which can lead to incomplete or misleading views of learners’ knowledge” (p. 221). The present study aimed at using multi-faceted instruments which covered a broad range of concepts within additive reasoning. Furthermore, Johnson and Christensen (2014) mention the problematic nature of using single test scores and promote using a range of data sources. The importance of multi-faceted assessment has been discussed; thus alongside tests used to measure fluency, diagnostic assessments (Table 3.1) were used to measure various concepts within additive reasoning. These will be discussed respectively.

Table 3.1

Data collection tools used with construct operationalised

Data collection tool	Construct operationalised	Research question
Fluency test	Fluency of addition and subtraction facts	One
Self-report tool	Fluency of addition and subtraction facts	One Five
Diagnostic questions	Understanding of additive reasoning	Two
Empty box test	Understanding of additive reasoning	Two
Conceptually ordered fluency (COF) test	Fluency of addition and subtraction facts	One and two

	Understanding of additive reasoning	
Semi-structured interviews	Student perception of the intervention Student perception of manipulatives Understanding of additive reasoning	Three Two Four

3.5 Fluency data collection method

Research question 1 explores whether the use of structured manipulatives improves the fluency of participants' recall of addition and subtraction facts. Within the literature a range of data collection methods is used to measure fluency and these were used to inform tools for the present study. Baroody (1999) used reaction time as a measure of addition and subtraction fact knowledge. This data collection method ensures that the data are sensitive to participants counting or using inefficient methods of calculation. Due to time and resource restraints, this data collection method was not used in the current study. However, the study used a method to calculate the number of correct responses in a given time to address the concerns of Baroody (1999) and the use of counting. The fluency test was based on a study examining fluency with multiplication (Nelson, Burns, Kanive, & Ysseldyke, 2013). In the current study, participants were provided with a test consisting of 60 questions (Appendix A). These were presented in three columns of 20 on A4 paper. Thirty of these questions were addition questions and 30 questions were subtraction questions. All of the questions involved whole numbers and all answers were whole numbers above zero. The questions required the participants to bridge 10 (e.g., 8 + 5 or 13 – 6). The questions alternated from addition to subtraction. Standardised tests of mathematical fluency such as the Woodcock-Johnson test also use a mixing of addition and subtraction (Starkey & McCandliss, 2014). Participants were instructed to answer questions by moving down the columns but could skip any questions that they were unsure of. Work in this field used questions of comparable difficulty, albeit with a younger sample (Baroody et al., 1983). All participants answered these questions in silence without the use of any manipulatives. After two minutes, they were instructed to stop answering the questions. The same test was provided at pre-test and post-test. Furthermore, a self-report question was asked after the fluency test.

3.6 Self-report after the fluency test

Research question 1 focuses on whether structured manipulatives improve fluency when recalling addition and subtraction facts. Research question 5 explores whether counting procedures are used frequently in this sample. These two questions are linked as it may be possible to perform well on the fluency test instrument using counting procedures. Moreover, the researcher was aware of a possible lack of validity using the same fluency test before and after the intervention. Participants were asked to self-report how they had answered questions after completing the fluency test instrument at pre-test and post-test. A self-report question allows the researcher to gain an insight into students' mathematical processes and strategies while also attempting to increase the validity of any findings (Johnson, 2014). Previous research in this field also used self-reports to gain an insight into participants' conceptual understanding (Baroody et al., 1983; Canobi et al., 1998). Baroody (1999) states the importance of knowing how students answered addition questions. In the current study, participants were provided with the written, open-ended question, "How did you solve these?" An open-ended question was chosen as it provided the greatest opportunity to gain insight into the mathematical thinking and processes of participants. Moreover, it was important that participants responded with language that made sense to them when describing their mathematical thinking rather than be constrained by the language of the researcher. Previous work in this field used open-ended self-report questions to collect data on learners' strategies (Canobi et al., 1998). Answers were coded inductively from the responses given by participants during the pre-test. The codes identified were: *used a counting strategy, memory, used an arithmetic strategy such as doubling, and miscellaneous*. The researcher checked the responses at post-test to ensure that the codes used at pre-test adequately reflected responses at post-test.

3.7 The rationale and development of diagnostic questions

After completing the fluency test and self-report, participants answered a range of diagnostic questions. This is supported by Canobi et al. (1998) who concluded it was meaningful to explore participants' explanation of additive concepts. Furthermore, previous work has highlighted the importance of using multi-faceted assessments (Prather & Alibali, 2009). The present study uses a range of data collection tools to operationalise addition reasoning; diagnostic questions are one of these tools.

3.8 Research on multi-faceted assessments to operationalise additive reasoning

Prather and Alibali (2009) discuss the role of context on participants' performance on verbal, concrete, abstract and symbolic arithmetic principles. Specifically, verbal context refers to the story context of a word problem, concrete context refers to the use of manipulatives, abstract context might involve the use of pronumerals, and symbolic use purely of numerals. When designing the diagnostic questions for this study, consideration was given to the context of the questions. The current study tried to remove verbal context by decontextualising the diagnostic questions. As the research questions focussed on whether the structured manipulatives improved participants' fluency and additive reasoning, it was decided that concrete materials would not be available for the testing phases. The study was concerned with whether thinking and understanding could be shifted after the intervention. The use of manipulatives in the testing phases might not reflect any changes in LTM which, in line with CLT, was a goal of the study. Furthermore, research has discussed how manipulative use is not the goal but a tool to achieve more sophisticated thinking (Bonne, 2003).

Prather and Alibali (2009) highlight the importance of multi-faceted knowledge assessments when assessing arithmetic principles. This paragraph will briefly summarise the work on multi-faceted knowledge assessments and identify which assessment types were used in the current study and the rationale for these choices. Several types of knowledge assessment are discussed: applications of procedures, evaluation of procedures, evaluation of examples, justification of procedures, and explicit recognition. Explicit recognition involves agreeing or disagreeing with an abstract statement that exemplifies a principle. For example, *is it possible to add two numbers in any order? Do you agree or disagree?* Explicit recognition of a concept can be proxied by the application of a procedure. For example, observing a student solving 3 + 99 by counting on from 99 might reflect additive reasoning understanding. However, there is some debate regarding whether this specific procedure accurately reflects understanding of commutativity (Baroody, 1987). When evaluating examples, participants rate questions and answers that have been provided by others. When participants rate an example that breaks a specific principle lower than an example that does not break the principle, this suggests that the participant has knowledge of the principle. For example, 10 – 7 = 13 might be rated worse than 10 – 7 = 2 as it shows an increase after subtraction of whole numbers. Evaluation and justification of procedures are closely related.

Participants look at the examples of others' mathematical thinking and evaluate whether these examples of thinking are correct; for example, a + b can be used to solve b + a. Justifications of procedures involves identifying the principle adhered to or broken; for example, you cannot subtract in any order (Prather & Alibali, 2009). The majority of diagnostic questions (Table 3.2) used in this study use evaluation and justification of procedures. One application of procedures question was created to assess whether participants could recall related subtraction and addition facts.

The rationale for using explanation and justification of procedures is explained by considering the disadvantages of other question types. First, application of procedures may require close observation of each participant. The intervention was completed outside of school and the time with participants was limited. Furthermore, questions categorised as application of procedures may involve some calculation. In the main, this study tried to operationalise fluency and additive reasoning separately (Section 3.10). The researcher felt that using explicit recognition questions from Prather and Alibali (2009) would not provide reliable enough data. If a participant was asked to agree or disagree with a statement, there is a 50% chance of guessing the correct answer. Therefore, the majority of diagnostic questions involved evaluating and justifying procedures, allowing for richer and more valid data.

The diagnostic questions used in this study were written so that calculation would not be required or possible and this is mirrored in the literature (Canobi, 2005). The numbers included within the diagnostic questions were more challenging than the fluency test. Three-digit, four-digit and decimal numbers were used. Also, the questions were written so that they would be challenging enough that retrieval strategies would be unlikely (Ding & Auxter, 2017; Gilmore et al., 2018). The rationale for this was to ensure that the diagnostic questions operationalised additive reasoning. This was a concern as the literature has identified that some students correctly answered inverse questions without using reasoning knowledge (Ding & Auxter, 2017). Participants completed the diagnostic test in silence without manipulatives; all questions were provided to participants on a test paper. These were also read to all participants in order to minimise the confounding influence of reading ability.

The literature informed the diagnostic test tools. Questions were written with consideration of Prather and Alibali's (2009) knowledge assessment types. Research in this field has operationalised additive reasoning in a range of ways that overlap the work of Prather and Alibali (2009). Baroody (1999) asked participants whether one addition or subtraction expression could help to solve a related addition or subtraction question; after this, participants were asked to justify their answer. A range of studies using a similar method has been completed (Cowan et al., 2011; Ding & Auxter, 2017; Dowker, 2014; Gilmore & Spelke, 2008). Canobi (2005) examined additive reasoning in a population of seven to nine year-olds, albeit using interviews rather than written responses. Participants had to evaluate whether a puppet needed to use counters to solve a calculation after the puppet had solved a related question. After an addition calculation, participants were provided with a commutative addition question and subtraction inverse question, 25 + 24 = 49 followed by 24 + 25 = 49 and 49 – 25. After a subtraction problem, an addition inverse and subtraction complement problem were provided, 64 – 21 = 43, followed by 21 + 43 and 64 – 43 = 21. Participants were required to decide whether the puppet could use the previous calculation to solve the problem. Canobi (2005) labelled these sub-concepts of additive reasoning: commutativity, subtraction inverse, additive inverse, and subtraction complement. These sub-concepts were used when creating the questions in the current study (Table 3.2).

Some of the questions included in the current study do not follow Canobi (2005). Dowker (2014) discusses how some errors are the result of over extension of an arithmetic principle, such as treating subtraction as commutative. The researcher felt that it was important that participants could discriminate where additive relationships exist and where they do not. Research in this field used similar techniques. Ching and Nunes (2017a) used questions to control for a response bias e.g., does 5 + 3 = 8 help you to solve 5 – 3? The misapplication of a principle suggests that it is not fully understood. Research has highlighted concerns that learners may use shortcuts without mastery of a principle (Baroody et al., 1983; Canobi, 2005). Therefore, in the current study, one diagnostic question asked participants to judge whether subtraction was commutative. Another question asked participants to judge whether a subtraction question could be answered after an addition question where the numbers remain constant but the operation changes, a + b and a – b. Question 3 (Table 3.2) uses application of procedures not explanation and justification (Prather & Alibali, 2009). The ability to derive three related facts from a given fact is a defining construct of additive reasoning (Ching & Nunes, 2017b). Therefore, participants in this study were asked to provide other related number facts when given an addition fact. The numbers used in this question were selected to discourage or inhibit the use of calculation.

In summary, the literature informed the design of the diagnostic questions (Appendix D). It was felt that evaluation and justification of procedures questions (Prather & Alibali, 2009) provided the most reliable data required to answer research question two. For all but one of the questions, participants were provided with an example of another person's mathematical thinking and were asked to evaluate whether that knowledge could be applied elsewhere. Participants were then asked to provide a written justification. Previous research in this field also provided participants with the opportunity to explain answers (Canobi, 2005; Ding & Auxter, 2017). However, it is important to note other work has suggested that asking students to explain strategies might encourage more flexible responses (Gilmore & Spelke, 2008). This might be considered a limitation of this data collection method.

Table 3.2

Diagnostic questions used with links to literature

Question	Questions	Additive reasoning focus	Assessment type (Prather & Alibali, 2009)	Theoretical basis
1	If Mark knows that 234 + 574 = 808, can he use this to solve the problem below? ☐ — 234 = 574 Yes ☐ No ☐ Not sure ☐ Explain your answer	Find a missing minuend when given the corresponding addition fact	Evaluation and justification of procedures	Subtraction inverse (Canobi, 2005)
2	If Mark knows that 1 – 0.12 = 0.88, can he use this to solve the problem below? 0.12 + ☐ = 1 Yes ☐ No ☐ Not sure ☐ Explain your answer	Identify a missing addend when provided with related subtraction fact	Evaluation and justification of procedures	Additive inverse (Canobi, 2005)
3	If Mark knows, 398 + 777 = 1175, what other number sentences can he make with these three numbers?	Identify three related facts from one addition fact	Application	The definition of additive reasoning (Ching & Nunes, 2017b)
4	If Mark knows 177 + 383 = 560 can he use this to solve 383 + 177? Yes ☐ No ☐ Not sure ☐ Explain your answer	Understand the commutative law of addition	Evaluation and justification of procedures	Commutativity (Canobi, 2005)
5	If Mark knows 177 + 383 = 560 can he use this to solve 383 – 177? Yes ☐ No ☐ Not sure ☐ Explain your answer	Understanding the process of, and difference between, addition and subtraction.	Evaluation and justification of procedures	Over extension/ misapplication of principles (Dowker, 2014)
6	If Mark knows the answer to 577 – 434, can he also solve 434 – 577? Explain your answer Yes ☐ No ☐ Not sure ☐	Understand that subtraction is not commutative	Evaluation and justification of procedures	Over extension/ misapplication of principles (Dowker, 2014)

3.9 Empty box test

After completing the diagnostic questions, participants attempted the empty box test (Figure 3.1). The empty box test was designed to operationalise additive reasoning. Participants were required to place the correct number in the empty box (Appendix B) in an addition or subtraction calculation. The position of the empty box varied from question to question but was never located after the equals sign to distinguish between these types of questions and questions that were asked in the fluency tests. Similar questions to these were used as part of the National Education Monitoring Project and were identified as challenging for approximately one third of year 8 students (Crooks, 2010). All calculations used only whole numbers and all questions involved bridging 10. The empty box test was presented in a similar format to the fluency test instrument; however, only 29 questions were provided. During piloting, performance on this test indicated fewer completed questions in comparison to the fluency test. The researcher was cognisant that participants should feel positive about the current study and therefore reduced the number of questions on the page. They were given two minutes to complete as many questions as possible. Participants completed this test in silence and were not provided with manipulatives. As with the fluency test, addition and subtraction questions alternated. The rationale for this data collection method will be explained alongside the rationale for the conceptually ordered fluency test (COF) (Section 3.10).

$\square - 8 = 5$
$5 + \square = 16$
$12 - \square = 4$
$\square + 9 = 16$

Figure 3.1. Empty box test example

3.10 The conceptually ordered fluency test

After the empty box test, participants completed the COF test. This test was presented in a similar fashion to the fluency test with questions of the same difficulty. However, questions were ordered so that related addition and subtraction facts were grouped together. Questions did not alternate between addition and subtraction. Within each set of four questions, two subtraction questions and two addition questions were presented. The four questions contain the related addition and subtraction facts (a + b = c, b + a = c, c – a = b, c – b = a) in various orders. Participants completed as many questions as they could in silence for two minutes and were not provided with any manipulatives. They were encouraged to answer questions in order by working down the columns but were allowed to skip any questions they found too challenging. Previous work in this field has measured conceptual understanding in the field of additive reasoning by providing participants with related calculations (Canobi et al., 1998; Gilmore & Spelke, 2008).

The literature informed the use of COF and empty box tests. Canobi (2009) highlights the challenge of extracting conceptual understanding of additive reasoning from participants as learners may not be able to verbalise their understanding. Torbeyns et al. (2016) used non-verbal data measures when investigating the complement principle. This study used a looking back task where a previous question at times provided the answer to a target question; this research called for the increased use of non-verbal data when measuring the complement principle. In light of this research, diagnostic questions alone may not adequately operationalise additive reasoning; therefore, the empty box and COF test aim is to operationalise understanding of additive reasoning using a task which does not require verbal explanation.

6 + 9 =
9 + 6 =
15 — 9 =
15 – 6 =
17 – 9 =
17 – 8 =
9 + 8 =
8 + 9 =

Figure 3.2. Conceptually ordered fluency test questions example

The fluency, empty box and COF tests constituted the pre and post-tests. Although these tests were designed to operationalise different concepts, they are very similar. To reduce testing effects, the researcher separated these during the data collection phases. As the fluency and COF tests are most similar, these were placed at the start and end of the data collection process respectively.

3.11 Semi-structured interviews

The aim of the interviews (Appendix E) was to examine perceptions of the intervention, manipulatives, and changes in mathematical understanding. This provides data for research questions 3 and 4. Participants were also asked one verbal diagnostic question which was similar to an evaluation of procedure diagnostic question. However, in line with Johnson (2014) who discusses the importance of sensitivity when conducting an interview, a decision was made on whether to ask a further diagnostic question based on each individual participant's responses and demeanour during the interview; for example, should a participant respond negatively to all the interview questions, leading the interviewer to feel that a further mathematical question would be damaging, then the diagnostic question was not asked. This was particularly important considering the maximum variation sampling used for the interviews.

The rationale for including a verbal diagnostic question to focus on research question 2 is that it allowed for the use of the responsive nature of semi-structured interviews in comparison to the less responsive diagnostic questions in the pre-test/post-test (Creswell, 2014). Furthermore, the research questions focus on the use of structured manipulatives. It was felt that a verbal diagnostic question afforded the opportunity to examine how participants used and perceived the manipulatives more closely, adding richness to the study. Participants were provided with manipulatives during the semi-structured interview.

This study used the interview guide approach to collect qualitative data (Johnson & Christensen, 2014). This approach is described as using an interview protocol (Appendix E) to explore specific questions or issues. The approach allows the interviewer to change the wording and order of questions, and to ensure that the content of the interview provides data relating to the research questions (Johnson & Christensen, 2014). This method of interviewing was chosen as it allowed the researcher the flexibility to explore the research questions using language familiar to the student. Another strength of this data collection method is the conversational nature of the interview (Johnson & Christensen, 2014). Many of the students in this study are relatively young and lack confidence and experience when articulating their learning goals and needs. Ethically, it was vital that participants were at ease and, from a research viewpoint, it was felt that the data would be more trustworthy if the students felt comfortable. However, Johnson and Christensen (2014) highlight some issues with this approach. Specifically, the interviews can differ considerably and this reduces their comparability. This is a possible limitation. As maximum variation sampling was used, this interview method allows flexibility for the researcher to adapt the interview wording and order for each participant where necessary whilst still collecting data aimed at answering the research questions. The maximum variation sampling selects a more heterogenous sample from the original participants and it was felt that flexibility was key during qualitative data collection to explore the research questions fully.

Creswell (2014) highlights the importance of using an interview protocol. This tool was used to provide structure and consistency to the interview and to ensure that appropriate data were collected to answer the research questions (Appendix E). The interview protocol used in this study follows the guide produced by Creswell (2014). All interviews began with the interviewer thanking the participant and informing them that the interview was about the mathematics intervention. Participants were told that they could stop the interview at any point. Also, participants were informed that there were no correct or incorrect answers and that they would not get into trouble for answers provided in an attempt to build trust (Johnson, 2014). The interviewer/researcher was a teacher at the school of the participants and it was important that any power relations were diminished. The interviewer made verbatim notes during the interview. These verbatim notes were read back to the participant to check for accuracy.

Johnson (2014) discusses the importance of impartiality when conducting interviews and to ensure reactions are neither positive nor negative. The interviewer was conscious of maintaining this impartiality during the interview. The wording of questions was intentionally neutral (Appendix E). Furthermore, questions were worded to elicit rich data to answer the research questions. Should a participant provide ambiguous answers, probes from Johnson (2014) were also used to clarify answers or explore a participant's response further (e.g., What do you mean?).

Table 3.3

Interview questions for exploring students' perceptions

Interview question	Rationale/purpose
How did you feel about the intervention?	To build trust and legitimise the voice of the interviewee
Can you explain why you felt RESPONSE A about the intervention?	To explore perceptions of the intervention
What effect do you think the intervention has had on your maths?	To explore participants' perceptions of the intervention and their learning
(Point to manipulatives) Can you talk to me about these?	To explore participants' perceptions of the manipulatives

In summary, a range of data collection tools was utilised in this study. Participants completed a fluency test to operationalise recall of basic addition and subtraction facts. After completing the fluency test, participants were asked to self-report how they answered the questions. Diagnostic questions were presented next and these were written with the aim of operationalising a range of additive reasoning sub-concepts. Next, the empty box test provided a range of addition and subtraction questions with an empty box placed before the equals sign. A conceptually ordered fluency test (COF) provided a range of addition and subtraction questions. Questions were specifically ordered based so that each group of four questions were from a fact family. The COF and empty box tests were designed to operationalise additive reasoning without relying on participants' verbal or written responses. All of these tests were completed before the intervention and again after the intervention. Upon the completion of the post-tests, four participants were invited to take part in semi-structured interviews and were purposefully selected using maximum variation sampling. The semi-structured interviews operationalise perceptions of the intervention and manipulatives, but also include one diagnostic question to utilise the responsive affordance of an interview.

3.12 The current study context

The study was carried out in an intermediate school in the South Island of New Zealand where the researcher was a teacher. The school has approximately 500 students and 20% of these identify as Maori. The school was a convenience sample chosen to enable easy access to participants who met the criteria for the study. The sample is small and the researcher has limited access to participants suitable for the study.

Year 7 and 8 students were selected as a focus for the study as these year groups fall within an age range where underachievement in mathematics persists in New Zealand (Chamberlain, 2013) and at a crucial time for influencing their future mathematical achievement and success before joining college. Lee (2010) identified the diminishing gains in mathematical achievement from primary school to middle school, and from middle school to secondary school. This suggests that finding effective means of acceleration at years 7 and 8 is crucial.

After gaining ethical consent from the Victoria University of Wellington Human Ethics Committee, the researcher wrote to the principal of the school selected for this study, informing him of the study. Once the principal had given permission for the study to take place, the researcher wrote to other teachers at the school asking them to nominate possible participants. Participants were to be fluent English speakers, not have a diagnosed special need, and be working at level two or early level three of the New Zealand curriculum in mathematics. Teachers were also asked not to nominate students who might experience harm from the intervention. After nomination, students and parents were provided with separate information and consent forms (Appendix F). Eleven participants agreed to participate in the intervention.

3.13 Intervention outline

The intervention took place once a week for seven consecutive weeks. Sessions were conducted after the school day and each session lasted approximately 45 minutes and took place in a classroom of the school that participants attended.

Participants were invited to sit where they wanted and they could choose to sit with peers or alone. Both types of structured materials (Numicon and Cuisenaire rods) were readily available for participants to use. Sufficient supplies of these manipulatives were available so that participants could select the manipulative they preferred or use both types. Participants were also provided with writing equipment.

A key aspect of the intervention was that sessions were responsive to the needs of the participants. Therefore, the seven intervention sessions were not planned before the intervention. The previous session informed the content of the next session. Opportunities for discussion and explanation were seen as teaching moments. However, aspects of each intervention session were consistent. Participants were encouraged to familiarise themselves with the equipment during each session. Furthermore, they were encouraged to orient the manipulatives to model the actions of the addition and subtraction questions. Participants were only encouraged to use the manipulatives if they had made a mistake or were observed counting by the researcher. It is possible to use the Numicon shapes as tools for counting strategies. If the researcher observed this, participants were encouraged to orient the shapes against a ten and use visual strategies to solve the problem. Furthermore, they were taught using terms that help to generalise additive relationships: part + part = whole and whole – part = part. Baroody (1999) suggests highlighting the parts and whole to teach the complement principle, citing personal communication with Fuson, "she conjectured that noticing the relation between such related subtraction combinations (Whole – Part 1 = Part 2 and Whole – Part 2 = Part 1) might be the critical discovery in understanding the complement principle" (p. 170).

Another consistent feature of intervention sessions was the teaching of both fluency and additive reasoning. At the beginning of each session, for approximately ten minutes, participants were given an opportunity to practise addition and subtraction facts. These questions were of the same difficulty as the questions in the pre-test and post-test (Appendix A). Participants were encouraged to use the manipulatives to help calculate or check the answers to questions. After practising fluency, the intervention moved on to practising an area of additive reasoning. This used a variety of techniques such as empty box questions, and ordered sets of calculations ($a + b = c$, $c – b = a$).

A consequence of the responsive nature of the intervention was that other areas of mathematics were discussed and taught alongside additive reasoning. Algebraic notation was also used to help learners to generalise this relationship, $a + b = c$ and, $c - b = a$. Furthermore, when teaching participants about the commutative nature of addition, and that subtraction was not commutative, participants increased their knowledge of negative numbers.

3.14 Analysis of fluency/empty box/conceptually ordered fluency tests

As the study design collected data in a range of ways, a range of tools was used to complete the analysis. Analysis of each data collection method will be discussed respectively.

All three tests (fluency, empty box, COF) were measured using the same scale. Previous research in the field of fluency used the measure *mean correct digit per minute* (DCM) to measure fluency (Burns, Codding, Boice, & Lukito, 2010). For a response such as

7 + 8 = 15, participants would score two points as both digits are correct and in the correct position. If the participant had answered the same question with '51' this would result in a score of 0 as the digits are not in the correct position. An answer of '16' would score one point for the position of the digit one. Descriptive statistics were collated for each test at pre-test and post-test respectively. Inferential statistics (t-tests) were used to compare each test at pre-test and post-test because, although these tests are measured using the same scale, it was expected that they will operationalise different items. Correlation coefficients were obtained to gain an insight into participant performance on the three tests at both time points, and to examine whether the three tests operationalise different constructs. After completing the fluency test, participants were asked to self-report their calculation strategies (Section 3.6). These qualitative data were coded inductively. Some statements were double coded. This reflects the research that learners often use a range of strategies to calculate (Siegler, 1987).

3.15 Analysis of diagnostic questions

The majority of the diagnostic questions required participants to answer 'yes', 'no', or 'not sure' and then to provide a justification for the chosen answer (Appendix D). Answers provided by participants were coded using *a priori* codes. The rationale for the use of *a priori* codes links to the construct being operationalised in research question 2: additive reasoning. Each diagnostic question operationalises a different area or concept within additive reasoning. *A priori* codes allow data to be enumerated across the different concepts spanning additive reasoning. The *a priori* codes (Table 3.3) that were used are similar to those used by Ding and Auxter (2017). Further, this affords analysis of whether the intervention is more effective at developing some areas of additive reasoning than others (Canobi, 2005). Despite the use of *a priori* codes, the researcher was prepared to change codes and use inductive coding if the *a priori* codes did not fit responses from students. The data analysis completed with diagnostic questions required some decision making regarding whether a participant's response is mathematically acceptable. Ding and Auxter (2017) argue that it is important to be aware of answers that are correct but do not have a justification as this may not indicate a lack of understanding. The codes and categories used reflect this consideration.

Table 3.4

A priori codes used for diagnostic questions

Code	Category
Not sure ticked	Incorrect
Incorrect answer	Incorrect
Correct box tick but incorrect explanation	Incorrect
Correct box ticked but no explanation or participant states they cannot explain	Correct box ticked but no explanation
Correct box ticked with mathematically acceptable justification	Correct

One diagnostic question was coded differently. Participants were asked to provide related facts when provided with one addition fact. They were asked to use the same three numbers in the original equation. Again, the researcher used *a priori* codes (Table 3.5) when analysing these data. It is important that the codes reflect possible misconceptions that participants have; for example, a participant including the equations $a - b$, and $b - a$, would be coded as *Equations provided (some incorrect and some correct).* This suggests that the participant has some understanding of the relationship between addition and subtraction but does not yet understand that subtraction is not commutative. As well as allowing for identification of misconceptions, the codes afford a way of measuring the developing understanding of principles, laws and relationships. Inductive coding was a consideration if these *a priori* codes were not suitable.

Table 3.5

Codes used for analysing related fact diagnostic question

Code
No equations provided
Only incorrect equations provided
Equations provided (some incorrect and some correct)
One or more of the correct equations provided (no incorrect equations)
All three correct equations provided (no incorrect equations)

3.16 Analysis of the semi-structured interviews

Coding of the data was completed inductively as this research aimed to augment the body of literature on additive reasoning and fluency in year 7 and 8 students rather than use an existing schema (Johnson, 2014). The coding of these data used a phenomenological approach, examining how the four interviewees experienced the intervention. In order to get participants to relive the experience, the semi-structured interviews were designed with questions to prompt participants with the intervention to help remind them of it (Johnson, 2014). The essence of participants' experience was reported with the questions were asked using verbatim quotes.

3.17 Validity and trustworthiness

Johnson (2014) describes validity as, "The accuracy of the inferences, interpretations, or actions made on the basis of a test score" (p. 172). Internal validity can be defined as, "The ability to infer that causal relationship exists between two variables" (Johnson & Christensen, 2014, p. 281). Single-group designs, such as the one used in this research, present threats to internal validity as any changes in responses gained from pre-test and post-test could be the result of a range of confounding variables rather than the intervention. History can be defined as any events that happen between the pre-test and post-test that might result in changes in the dependent variable (Johnson, 2014). When examining this study, history needs to be considered as the participants came from different classes in the school. It was not possible to control for the type, amount, or quality of mathematics being taught in participants' respective classrooms during the intervention. Furthermore, being selected for the intervention might have triggered participants to focus more on mathematics at home. Although these threats to internal validity need to be considered, Johnson (2014) does state that history is most problematic during lengthy interventions and this intervention lasted only seven weeks, so it might be argued that history is less problematic in the present study.

Maturation is another threat to internal validity that needs to be considered. Johnson (2014) defines maturation as "mental changes that may occur in individuals over time" (p. 285). Any changes from pre-test to post-test may be the result of maturation rather than the intervention. Again, it might be argued that threats from maturation are less problematic due to the short nature of this intervention.

Testing is the phenomenon of participants' improved results on a test occurring because they have previously completed the same test. Again, the study design used in this research is vulnerable to this threat due to the same test being used before and after the intervention. Johnson (2014) discusses how familiarity with the content of a test can boost performance. This study tried to mitigate against this somewhat by not providing participants with feedback on their performance after the pre-test, and students not having access to the test during the intervention. However, it is possible that testing is a threat to internal validity.

External validity is defined as, "the extent to which the results of a study can be generalised to, and across populations" Johnson (2014, p. 291). The sample selected for this study was not randomly selected from the target population but was a convenience sample nominated by class teachers. Therefore, any generalisations from the study will be limited, but may provide direction for future research. Furthermore, it will not be possible to generalise the findings to other settings as it was difficult to control for effects of the teacher in this study. Pedagogical content knowledge, subject knowledge, relational aspects of teaching, and the ability to provide effective feedback may vary depending on who is implementing the intervention (Hattie & Timperley, 2007; Shulman, 1986). As such, the study lacks ecological validity which can be defined as the ability to generalise results across studies (Johnson, 2014). This links closely with the concept of treatment variation validity, "the ability to generalize the results across variations" (Johnson, 2014, p. 294). As stated previously, varying the teacher implementing the research may lead to different outcomes. When considering treatment variation validity, it would also be salient to consider the number of participants in the group. There was a considerable amount of student to student, and student to teacher talk within the intervention, and changing the size of the group may impact on this negatively or positively. The intervention was very much in response to the participants; the structured manipulatives only formed part of the treatment. It is not possible to completely separate the effectiveness of the structured manipulatives from other aspects of the intervention. Future research might use a control group to examine this more closely. Another consideration could be the behaviour of participants. Participants who are more or less likely to share their mathematical thinking may alter results. This is especially salient as the intervention was responsive to comments made by participants. Furthermore, this study took place after the school day. We cannot generalise findings from the study to interventions completed during the school day due to students being more tired at the end of the school day and the possible stigma that low-achieving peers may feel by being involved in extra mathematics learning time.

Validity or trustworthiness is also applicable to qualitative research and can be defined as research that is "plausible, credible, trustworthy, and therefore defensible" (Johnson, 2014, p. 299). Johnson highlights a range of strategies to increase validity such as a critical friend providing feedback to the researcher. This strategy was used during the research as the researcher's supervisor provided regular feedback.

The data from the interviews were coded and verbatim quotes were used during the analysis. Furthermore, characteristic to the phenomenological approach, bracketing occurred before analysis of the interviews. This was an attempt at reducing researcher bias through the conscious removal of preconceived ideas (Johnson, 2014). Moreover, maximum variation sampling was used when selecting participants for the interview phase. Within this, there were some cases that could be defined as negative cases (i.e., students who made the least amount of progress on the fluency measure). Johnson (2014) states that negative case sampling is a useful strategy in mitigating for research bias as these cases can disconfirm beliefs that the researcher holds. Finally, the researcher attempted to increase trustworthiness through a study design that maximised opportunities for triangulation. Table 3.1 shows how multiple data collection methods were obtained in an attempt to corroborate one another and provide rich data.

Chapter 4

The description of data analysis will be carried out in the same order as the data collection methods were administered as is appropriate for presenting results of a sequential explanatory design study. One participant did not complete the intervention; the data presented are from the remaining ten participants only.

4.1 Descriptive statistics show increase in mean correct digits between pre-test and post-test for all test instruments

Participants completed six tests (fluency, empty box, and COF) before and after the intervention. Across all the three tests, participants' mean correct digits per minute (MCD) increased from pre-test to post-test (Table 4.1). Measures of standard deviation showed that the spread of participants' scores also increased from pre-test to post-test for all test instruments (Table 4.1). The data were examined for normal distribution. Skewness and kurtosis were within the range of -2 and 2 for all data sets. Further to this, a Shapiro-Wilk test for normal distribution suggests that all data sets are normally distributed, $p>.05$. However, it is important to take into account the small sample size (10) used in this study when considering normal distribution.

Table 4.1

Mean (M) and Standard deviations (SD) from quantitative tests (Mean digits correct per minute)

Fluency Pre-test	Fluency Post-test	Empty box Pre-test	Empty box Post-test	Conceptually ordered Pre-test	Conceptually ordered Post-test
$M = 9.8$	$M = 16.4$	$M = 3.1$	$M = 5.4$	$M = 8.8$	$M = 13.4$
$SD = 3.3$	$SD = 5.8$	$SD = 2.7$	$SD = 3.5$	$SD = 4.3$	$SD = 5.2$

Note: N = 10, means and standard deviations to 1 decimal place.

4.2 Inferential statistical analysis demonstrating significant improvements from pre-test to post-test

Three dependent sample t-tests were used to compare the differences between MCD at pre-test and post-test for all three test instruments respectively. Bonferroni's correction was used to modify alpha levels to $p = 0.017$.

Despite the small sample size, it appears that the mean scores of all tests are significantly different at pre-test and post-test (Table 4.1). There was a significant difference in the scores for fluency at pre-test (M=9.8, SD=3.3) and post-test (M=16.4, SD=5.8); $t(9)=4.34$, $p = 0.0018$. There was also a significant difference in the scores for empty box at pre-test (M=3.05, SD=2.7) and post-test (M=5.35, SD=3.5); $t(9)=3.45$, $p = 0.007$. There was a significant difference in the scores from COF at pre-test (M=8.8, SD=4.3) compared to post-test (M=13.35, SD=5.2); $t(9)=3.47$, $p = 0.007$.

The differences between proponents of the drilling of procedural knowledge and the more conceptually driven flexible approach to arithmetic development have been discussed (Section 2.4). Research discussing the intertwined relationship between fluency and additive reasoning was shared (Section 2.12). Therefore, correlation coefficients were obtained from MCD scores on the fluency test and the empty box test at both pre-test and post-test. Pearson correlation coefficients were found for the fluency and empty box test at pre-test (0.67) and post-test (0.83). However, this improvement in correlations was not statistically significant. Before the intervention, the relationship between scores on the two test instruments was positive and moderately correlated. After the intervention, the scores from both tests were more closely correlated but this change was not significant. Future research might examine these tools further to gain a deeper understanding of the intertwined nature of conceptual and procedural understanding.

The moderate correlation of fluency and empty box at pre-test provides some evidence to suggest that participants were able to answer fluency questions (possibly using counting strategies) but this did not equate to performance on the empty box test. Although performance on the two tests was more strongly correlated at post-test, the change was not significant.

4.3 Inferential statistics highlighting that the test instruments used operationalised different constructs

The three test instruments were designed to operationalise similar but different constructs. This study aimed to augment the body of the literature on fluency and additive reasoning by combining a range of data collection methods (Prather & Alibali, 2009; Torbeyns et al., 2016). Therefore, it is pertinent to examine whether fluency, empty box, and COF tests did operationalise different constructs. The moderate correlation coefficients found at pre-test (Section 4.2) suggest that the fluency and empty box tests may operationalise different constructs. Inferential statistics were used to examine the differences in means between all three tests at pre-test in an attempt to strengthen the notion that the three test measures operationalise different constructs and inform any analysis of data collected from these tests. The decision to examine the means between tests at pre-test rather than post-test was justified by considering any possible improvements in scores due to the intervention, maturation or testing effects. Bonferroni's correction was used to modify alpha levels to $p = 0.017$ (3dp). There was a significant difference in the mean scores (Table 4.1) for fluency at pre-test ($M=9.8$, $SD=3.3$) and empty box at pre-test ($M=3.05$, $SD=2.7$); $t(9)=8.66$, $p = 0.00001$. There was not a significant difference in the scores for fluency at pre-test ($M=9.8$, $SD=3.3$) and COF at pre-test ($M=8.8$, $SD=4.3$); $t(9)=1.34$, $p = 0.21$. There was a significant difference in the scores for empty box at pre-test ($M=3.05$, $SD=2.7$) and COF at pre-test ($M=8.8$, $SD=4.3$); $t(9)=7.22$, $p = 0.00005$.

T-tests comparing the test instruments at pre-test show participants performed significantly worse on the empty box test than fluency and COF at pre-test. This result provides some weight to the assertion that the empty box test instrument operationalises a different construct to the other two tests. Added to this, the dependent samples t-test showed that mean performance on the fluency and COF instruments was not significantly different at pre-test. The lack of significant difference at this time may have several explanations. Firstly, there is a possibility that these tests may operationalise similar (or the same) constructs. Secondly, participants may have used the same strategies on both tests at pre-test. This is relevant because both tests used the same difficulty of question. It is also relevant to the current study which explores the possibility that counting strategies are used frequently in this sample. It is possible that individual differences in strategy use might result in the test operationalising different constructs, and this would be a limitation of the research.

A difference between the COF and fluency tests was the relationally ordered subtraction and addition questions in the COF test rather than randomly ordered questions in the fluency test. It was hypothesised that participants with more developed additive reasoning knowledge would perform better at this test as answers could be derived from previous questions. It might be expected that participants' scores would not differ significantly between the two test types at pre-test as participants may have been relying on similar strategies for both instruments. The results of the fluency and COF instruments were also analysed at post-test using a dependent sample t-test. There was a significant difference between the scores for fluency at post-test ($M=16.4$, $SD=5.8$) and COF at post-test ($M=13.35$, $SD=5.2$); $t(9)=2.79$, $p = 0.02$.

Participants performed worse on the COF test compared to fluency at post-test. This was an unexpected result. A closer look at the data shows that only one participant scored higher in the COF post-test compared to the fluency post-test. There are several possible explanations for this. It is possible that the intervention had little or no impact on additive reasoning; however, the significant increase in performance on the empty box test suggests that there were some gains in additive reasoning. It is possible that a flawed test design resulted in poorer performance on the COF test. The ordered nature of the COF test was not made explicit, and participants were given two minutes to complete the task. Because of this, it is possible that participants did not realise that the questions were relationally ordered. Therefore, it is possible that participants did not apply knowledge and strategies they had an understanding of due to a lack of awareness of how the task was organised. Another possibility is that individual understanding of additive principles was still developing and therefore participants may have been using the relationally ordered calculations, but this was slower than the strategies used in the fluency test. A limitation of this study is that a self-report question was not used at the end of the COF test to explore strategy use. A final consideration was that the COF was the final test completed in the post-test and comparing the performance on this test to the fluency test (which was completed first) might not be reliable as participants may have become fatigued as the data collection progressed.

4.4 What limitations exist in the quantitative phase of this research?

Participants showed significant improvement in all three test instruments; however, an increase in standard deviations for all instruments (Section 4.1) suggests that not all participants improved at the same rate. One explanation for this could be that despite the inclusion criteria, the group was not homogenous. During the intervention, the researcher noted that there appeared to be considerable variation in mathematical ability and understanding across the group. This was evident from the varying speeds that participants completed exercises during the intervention. Furthermore, the researcher noted there was considerable variability in the mathematical language that participants used. Participants were not nominated for the study if they had a diagnosed special need. It is possible that some members of the sample had an undiagnosed special need and made less progress in the intervention relative to the group. Future research might select participants more diagnostically (and from a wider population) in order to create greater homogeneity in the sample. The large variance within the small study sample makes generalisation of findings difficult.

The possibility that the improvements found in the quantitative tools of the study were due to a testing effect need to be considered. Some evidence for this was observed when results were collated; specifically, a greater number of questions were skipped in the post-test than in the pre-test. The significant improvement in mean scores might be attributed to students using the same counting strategies but being more strategic with their choice of questions. Future research might provide students with one question at a time to increase reliability. Students were informed that they could skip questions in the fluency and empty box tests. The rationale for this was to reduce any possible anxiety and discomfort caused by the testing phase; however, this affordance may confound results. Another possibility is that participants still required the manipulatives during the testing phase, and because they had not been reminded to use sophisticated strategies, they returned to counting strategies.

In conclusion, there was a significant improvement in the participants' performance on the test instruments from pre-test to post-test. Inferential statistics showed significant differences in the three tests at pre-test suggesting they operationalise different constructs. An unexpected result showed significantly worse performance on the COF test compared to the fluency test at post-test.

4.5 Analysis of diagnostic tests

Participants completed a range of diagnostic questions. In the quantitative phase of the research, significant improvements observed may have been the result of testing effects as the pre-test and post-test used the same questions. To explore the impact of the intervention further using qualitative data, diagnostic questions were included in both testing phases. Each question aimed to operationalise a different aspect of additive understanding (Table 3.2).

The results of these data were coded and then categorised to provide percentages of responses for each category at pre-test and post-test. The analysis provides insight into the specific changes in understanding during the intervention. There were two diagnostic questions coded differently. Firstly, participants were asked to self-report how they had answered the fluency questions. Secondly, participants were asked to derive addition and subtraction facts when given one fact.

4.6 Counting strategies self-reported by participants at both pre-test and post-test

After completing the fluency test, participants were asked to provide a written answer to the question, ‘How did you solve these?’ (Section 3.6). Participant responses provide further insight into the mathematical understanding and strategy use of the participants and the variance within the sample. At pre-test, the two miscellaneous comments were, ‘I used times tables’ and, ‘To make it easier, I swapped them around’. The first comment gives an indication of the mathematical understanding of this participant and the confusion of operations. The second comment suggests an understanding of the commutative law of addition; however, it is not clear from the participant’s response how switching 6 + 9 to 9 + 6, for example, helped the participant to solve the calculation. Therefore, this was coded as miscellaneous. It could be conjectured that switching addends might make ‘counting on’ a more efficient strategy as the participants will count from the largest addend (Baroody, 1987). However, without follow-up discussion with participants we cannot conjecture with confidence on a participant’s thoughts.

Some participants named counting strategies such as ‘count on’ or ‘count back’ while others simply stated that they ‘counted’. Some participants stated that they used their fingers, suggesting use of a counting strategy. On one participant’s pre-test, single tally marks showed further evidence of counting strategies. The researcher noted a large proportion of the participants using fingers to count during the pre-test. This adds weight to the hypothesis that participants in this sample use counting strategies to solve addition and subtraction equations.

It was hypothesised that from pre-test to post-test there would be a reduction in counting strategies used in the fluency and COF tests. The evidence (Table 4.2) does not fully support this. The same number of ‘counting’ responses were reported at pre-test and post-test. However, a greater number of responses were coded as, ‘memory’ or ‘arithmetic strategy’ than in the pre-test. This result suggests that some participants were beginning to identify that they were moving away from counting strategies while other participants persisted with counting strategies. This is supported by results (Table 4.1) which identified the increase in standard deviations from pre-test to post-test. Furthermore, the self-report comments from some students at post-test show a range of more sophisticated strategies being used. A lack of homogeneity within the sample was given as a possible explanation for this and Table 4.3 provides further evidence of considerable variation within the sample. Table 4.3 provides verbatim quotes of fluency self-reports at pre-test and post-test. These have been ordered so that the participant with the highest fluency post-test score is at the top of the table. It is noticeable that retrieval strategies appear to be used more by the students with higher fluency scores and, students with lower fluency scores rely on counting strategies.

Table 4.2

Number of self-reported responses of strategy use in the fluency test

	Counting strategy	Memory	Arithmetic strategy	Miscellaneous
Pre-test	5	2	1	2
Post-test	5	4	2	1

Note: responses could be coded in more than one way

Table 4.3 *Self-reported strategy choices in the fluency test ordered by performance*

Participant fluency score at post-test (MCD)	Self-report Pre-test	Self-report post-test
26.5	For 9 + 6 you do 10 + 6 = 16 -1 = 15	7 + 5 = 12 5 + 5 = 10 + 2
21	I counted some and some I knew	I just knew them
20.5	To make it easier I swapped it around	To make it easier I swapped it around
19.5	By doing x tables	For example 7 + 8 I knew that 4 + 4 was 8 then I added what need to be added
19	By counting with takeaway	I knew some off by heart and thinking of the block things and counting and taking away
14.5	Some I knew some I counted	I counted some of them
14	Counted them to get the right answer	I knew some Counting
10	Already knew them	Already knew some
10	With my fingers	I use my fingers to count
9	By using my fingers	I counted on my fingers

When examining the fluency test instruments, inferential statistics showed significantly different means from pre-test to post-test. Yet, the number of responses coded as ‘counting’ at both pre-test and post-test remained the same. One possibility could be that some participants improved their performance by becoming more efficient users of counting strategies. However, during the intervention phase, participants were specifically asked to try not to count and rather to use the manipulatives to help them solve the calculations. Moreover, participants were not taught or reminded of counting strategies during the intervention, although more developed counting strategies is a possible explanation for the improvement, Table 4.3 appears to suggest that counting strategies are being used by the participants with lower fluency. Aggregating the self-reports across the sample obscures some of the information. It is possible that the impact of the intervention on fluency and frequency of counting strategies varied across the sample. A limitation of the self-report tool is that it gives no information on how frequently participants counted or retrieved, nor on how strategies adapted were based on question type or difficulty (Siegler, 1987).

4.7 Analysis of diagnostic questions data collection phase: how the intervention impacted on understanding of additive reasoning

This section will analyse how participants’ performance on the diagnostic questions changed from pre-test to post-test. Further, it will describe how participants’ performances on different additive reasoning areas varied. Section 4.8 will provide a summary of results across all diagnostic questions. Subsequent sections will analyse each question type. Three categories of coded answers are reported. Incorrect responses are the proportion of responses coded as *Not sure, Incorrect answer or Correct box ticked with incorrect explanation.* Correct responses are only the responses coded *Correct with mathematically accurate explanation.* It is accepted that the participants in this study may have difficulty articulating their understanding and this difficulty may be characteristic of participants’ current levels of mathematical achievement (Torbeyns et al., 2016). Therefore, *correct answers without an explanation* will also be shared.

4.8 Summary of diagnostic questions showing an improvement in additive understanding

When examining responses to the diagnostic questions, it can be seen that 34% of correct answers were supplemented with a mathematically correct or acceptable explanation at post-test. In comparison, 8% of answers were coded this way at pre-test (Table 4.4). This suggests an improvement in accuracy and understanding of additive reasoning over the time of the study. At pre-test, 12% of participants gave a correct answer with either no explanation or noting that they could not explain their answer. This increased to 20% at post-test. These figures need to be interpreted with caution due to the lack of explanation, and therefore, the possibility of a participant randomly selecting the correct answer. A limitation of the diagnostic questions would be the inability to ask follow-up questions to clarify responses. However, the lack of nuance in this data collection method was partly addressed by a diagnostic question asked during semi-structured interviews (Section 4.23).

Table 4.4 shows that 80% of answers at pre-test were categorised as incorrect. At post-test, this reduced to 44%. These results suggest that the sample made some progress identifying and explaining key additive structures over the intervention. However, even at post-test, 44% of answers were categorised as incorrect. Furthermore, the averaging of percentages across question types obscures information. For example, the 8% of correct answers at pre-test were all from the same question: assessing understanding of the commutative law of addition.

Table 4.4

Percentage of answers across diagnostic question types that could be coded homogeneously

	Not sure	Incorrect answer	Correct box ticked but incorrect explanation	Correct box ticked but no response/I can't explain	Correct with mathematically accurate explanation
Pre-test	46%	26%	8%	12%	8%
Post-test	18%	22%	4%	20%	34%

4.9 Some participants showed a shift in ability to find a missing minuend when given the corresponding addition fact

Participants were asked whether it would be possible to find a missing minuend when provided with one of the two related addition facts. The two addends of the addition fact total the missing minuend. The subtrahend and difference are the same numbers used as addends in the given fact. During the pre-test, no students provided a correct answer with an accurate explanation (Table 4.5). At post-test, 20% of responses were coded as correct with an explanation. Further to this, 40% of students ticked the correct answer yet were unable to give an explanation compared to 20% at pre-test. At pre-test, 80% of answers were coded as incorrect with responses of: *Not sure, incorrect answer or correct box ticked with incorrect explanation*. Forty percent were coded the same way at post-test. A greater number of students were able to explain why it is possible to find a missing minuend from a related addition fact at post-test; however, this area of additive reasoning appeared particularly challenging to acquire and the majority of participants failed to answer this question.

Table 4.5

Finding a missing minuend (From 234 + 574 = 808 can you solve ? – 234 = 574)

	Not sure	Incorrect answer	Correct box ticked but incorrect explanation	Correct box ticked but no response/I can't explain	Correct with mathematically accurate explanation
Pre-test	30%	30%	20%	20%	0%
Post-test	20%	0%	20%	40%	20%

4.10 The intervention had limited impact on participants' ability to identify a missing addend when provided with a related subtraction fact

Participants were asked to find a missing addend when given a related subtraction fact. At pre-test, 80% of answers were incorrect (Table 4.6) and 0% of participants provided a correct answer with explanation. At post-test, this increased to only 10%. Overall, participants performed poorly when trying to identify a missing addend when given a related subtraction fact at both pre-test and post-test. It is possible that the use of decimals in the questions may have been a contributing factor; however, research discussed in Section 2.2 highlighted how manipulating an addition fact to find a minuend was challenging.

Table 4.6

Missing addend when provided with related subtraction fact (*If Mark knows that 1 - 0.12 = 0.88 what is 0.12 + ? = 1)*

	Not sure	Incorrect answer	Correct box ticked but incorrect explanation	Correct box ticked but no response/I can't explain	Correct with mathematicall y accurate explanation
Pre-test	60%	10%	10%	20%	0%
Post-test	50%	20%	0%	20%	10%

4.11 Participants showed an increase in their ability to identify related addition and subtraction facts

Participants were asked to provide related addition and subtraction facts when provided with one addition fact. The coding for this question was adapted to consider possible responses which included a misapplication of additive relations. For example, if I know 7 - 3 then I also know 3 - 7. A response like this would be coded as 'some correct facts combined with incorrect facts'. The researcher accepts that, for the example provided, knowledge of the fact 7 - 3 = 4 can be applied to 3 - 7 = - 4; however, no responses identified a negative number being derived this way. At pre-test, no participants attempted this question (Table 4.7). There was an improvement at post-test; only 10% of participants did not respond while 60% of participants provided the addition fact in its commutative form and the two related subtraction facts. Twenty percent of participants provided some of the possible facts while one participant (10%) responded with some correct and incorrect facts. This evidence suggests that participants at the beginning of this study did not have an understanding of the relationship between addition and subtraction. After the intervention, a greater number of participants were able to demonstrate an understanding of the relationship between addition and subtraction. Moreover, participants were able to identify this relationship with numbers that were challenging.

Table 4.7

Finding three related facts when given one addition fact (If Mark knows, 398 + 777 = 1175 what other number sentences can he make with these three numbers?)

	No answer	Only incorrect facts provided	Some correct facts combined with incorrect facts	Some available facts given (all correct	All available facts given
Pre-test	100%	0%	0%	0%	0%
Post-test	10%	0%	10%	20%	60%

4.12 A greater number of participants had an understanding of the commutative law of addition after the intervention

At pre-test, 40% of participants identified that a second addition fact could be derived from one addition fact and gave a correct mathematical explanation. This is surprising given that the previous question (given one addition fact what other number sentences could be made with these three numbers?) had 0% correct answers whereas 40% percent of participants could answer a commutative question correctly at pre-test. It is important to note the differences in question type. It is also possible that the difference in correct answers on these two question types is due to the assessment type (Prather & Alibali, 2009). Identifying linked addition and subtraction questions may be less cognitively challenging than deciding whether one calculation can help answer another and explaining why. The commutative question yielded the greatest improvement from pre-test to post-test in that 90% of participants could identify and explain use of the commutative law at post-test.

Table 4.8

Understanding commutative law of addition (If Mark knows, 177 + 383 = 560 can he use this to solve 383 + 177?)

	Not sure	Incorrect answer	Correct box ticked but incorrect explanation	Correct box ticked but no response/I can't explain	Correct with mathematically accurate explanation
Pre-test	30%	10%	0%	20%	40%
Post-test	0%	0%	0%	10%	90%

4.13 Participants still had limited understanding of the different operators when numbers remained the same

One of the challenging aspects of the symbolic notation of mathematics is that subtle changes in notation and order can be mean significantly different actions and answers (Haylock, 2010). The previous diagnostic question (Section 4.12) measured whether participants understood a + b = b + a. This diagnostic question asks whether there is a relationship between a + b and a - b. At pre-test, 70% of participants provided answers that were coded as *Not Sure;* the remaining 30% of answers were incorrect. At post-test, only 10% of answers were coded as *Not Sure* while 50% of answers were incorrect (Table 4.9). Of the incorrect answers ticked, only one provided an explanation, 'Yes, it's just turned around.' This suggests that participants are developing conceptual understanding of addition and subtraction yet misapplying this understanding to make incorrect generalisations. The majority of participants in this sample were unable to correctly answer this question.

Table 4.9

Understanding the process of, and difference between, addition and subtraction. Can we find 383 – 177 from 383 + 177?

	Not sure	Incorrect answer	Correct box ticked but incorrect explanation	Correct box ticked but no response/I can't explain	Correct with mathematically accurate explanation
Pre-test	70%	30%	0%	0%	0%
Post-test	10%	50%	0%	20%	20%

4.14 After the intervention, some participants still appeared unsure of whether subtraction is commutative

A large proportion of participants in this study were able to identify the commutative law of addition after the intervention (Table 4.9). This question aimed to test participants' understanding of subtraction. Participants were asked, *If Mark knows 577 - 434, can he use this to solve 434 - 577?* The researcher recognises that a - b = c and that b - a = - c where a is greater than b. Therefore, if a participant had ticked yes and linked their explanation to negative numbers in the above manner, this would be coded as correct. At pre-test, none of the participants were able to give a mathematically correct answer to this question. After the intervention, 30% of participants were able to give a correct answer and explanation while 40% gave an incorrect answer (Table 4.10).

Table 4.10
Understanding that subtraction is not commutative (If Mark knows, 577 - 434, can he use this to solve 434 - 577)

	Not sure	Incorrect answer	Correct box ticked but incorrect explanation	Correct box ticked but no response/I can't explain	Correct with mathematically accurate explanation
Pre-test	40%	50%	10%	0%	0%
Post-test	10%	40%	0%	20%	30%

4.15 Summary of results from the diagnostic questions used at pre-test and post-test

The diagnostic questions used in this study provide detail of the impact of the intervention. It appears that the intervention has made some improvement on the additive reasoning of some participants. This result is dependent on the concept of additive reasoning being tested. Participants' understanding of the commutative law of addition improved from pre-test to post-test. Furthermore, participants' ability to derive related addition and subtraction facts from one addition fact improved. The variation in the sample was once again evident as were individual differences in understanding, as a section of the cohort found it difficult to work flexibly with both operations, or incorrectly generalised a property of one operation to another (e.g., the commutative law of addition). This will be examined further in Section 5.2.

4.16 Limitations

The data collected from the diagnostic questions supplement the data collected from the quantitative phase. However, this collection method has some limitations. Although participants were able to ask clarification questions during the testing, this was in a group situation and participants may not have felt comfortable doing this in the presence of peers. Another limitation was the variation in numbers used in some of the questions. Numbers were chosen so that participants were unable to calculate with (or unlikely to calculate with) them in order to operationalise additive reasoning rather than ability to calculate. However, one of the question types used decimals, and another used numbers not bridging 1,000 and another used numbers bridging 1,000. It is possible that participants' answers were influenced by the numbers used in the equations as these may be numbers that participants do not have regular exposure to. Furthermore, even though a mathematical explanation was asked for, some participants did not provide explanations and some explanations required further clarification, leading to the researcher making subjective judgements when coding. Future research might use the same data collection method but verbally in a one-to-one setting. This might allow participants to ask for clarification of the diagnostic questions resulting in more accurate profiles of their understanding. A final phase of the experimental design was to conduct semi-structured interviews with four of the participants to provide greater insight into their perspective of the intervention.

4.17 Analysing the data collected from semi-structured interviews

Four participants were invited for semi-structured interviews. Scores on the fluency test were used to select participants using maximum variation sampling. The rationale for this was that the fluency test instrument had the largest standard deviation (Table 4.1) at post-test, and therefore the changes in scores were more evident for sampling. The aim of the semi-structured interviews was to gather information about participants' lived experience of the intervention. It was of particular interest to see whether this lived experience varied with mathematical achievement. Pseudonyms are used for the reporting of these data. Layla had the greatest fluency score (26.5 DCM) at post-test. Lucy had a score close to the mean at post-test (14 DCM where the mean = 16.4 DCM) and an improvement close to the mean improvement (6 DCM where the mean improvement was 6.6 DCM). Evie had a post-test fluency score below the mean (10 DCM where the mean was 16 DCM) and an improvement of 2 when the mean improvement was 6.6 DCM). Aiden had a post-test fluency score below the mean (10 DCM where the mean was 16 DCM) and an improvement of four when the mean was 6.6 DCM). Relative to the study, Aiden and Evie had lower mathematical achievement on the fluency instrument. Lucy displayed middling mathematical achievement and Layla had the highest level of mathematical achievement.

4.18 Participants felt that the intervention was positive

Lucy and Layla, who had middling and high achievement respectively, felt that the experience was positive.

> *Great!*
> *It was good.*

Aidan and Evie, who had lower mathematical achievement, gave responses that were slightly less enthusiastic, or placed the intervention into the context of a school day or considered the impact on their lives:

> *After school was a bit bad because it was after school.*
> *Nothing was bad.*

When asked to elaborate on the intervention, three of the four interviewees specifically mentioned the positive impact of the intervention on learning. This is especially pertinent as the interviewees had made varying levels of progress on the test measures. The participant who did not make a positive comment about learning had the lowest level of progress; they were unable to verbalise the effect that the intervention had on their learning:

> *We did extra maths and we learnt extra stuff.*
> *The learning was cool. It has helped.*
> *I learnt things that I didn't know how to do.*

4.19 Participants identify changes in their fluency strategies

Section 4.18 identified that most participants interviewed felt that the intervention had a positive impact on learning. All interviewees commented on how they felt they were less likely to count to calculate. The researcher did not mention counting or retrieval strategies during the interview, yet at least one of these was mentioned by all interviewees:

> *We learnt not to count.*
> *I'm better… because (before) I couldn't remember and had to use my fingers.*

I got better at not counting. I know things like 4 + 7 = 11 before I had to count. Before I used to add with my fingers. 9 + 6 = 15.

4.20 Participants' self-report transition from counting to knowing

Section 4.6 discussed that participants self-reported counting as a method of calculating at both pre-test and post-test in equal amounts. A common theme from the interviews was that interviewees report moving away from counting methods but still rely on this at some point:

I still use my fingers but not as much.
I know some facts.
Before I used to add with my fingers. I still do a bit, depends on the questions.
I got better at takeaways.

One comment from Layla (the highest achieving student) was particularly powerful in understanding how the intervention helped to develop her as a mathematician:

I don't use my fingers because I found better ways

4.21 Participants' attitudes towards the equipment used in the intervention

The majority of the four interviewees explained their preference for Numicon shapes rather than Cuisenaire rods. A common theme was that the Numicon shapes were easier to work with due to their numerical value being indicated by the number of holes in each shape:

I used these (Numicon shapes) to know which numbers are which. It's easier than the rods.
I used Numicon because I can see what the number is.
The equipment helped but the rods not as much because you can't see the numbers. I can pick up (picked up black Cuisenaire rod) but I don't know what it is.

Lucy, the student with middling progress and achievement, had a slightly different view of the Cuisenaire rods which was linked to her personal schema of the world and her interest in music:

I liked them both the same. I like using equipment. I said the rods were like the marimba because music and maths are connected.

One consideration for the data presented above is the relatively short length of the intervention. Although time was provided for participants to familiarise themselves with both types of equipment, perhaps a greater amount of time exploring and making sense of the manipulatives would have led to participants having more positive attitudes towards Cuisenaire rods. This is a consideration for future research in this area.

The essence of the experience for these four participants was that they felt positive about the equipment (especially the Numicon shapes) and a common theme was that the equipment was beneficial to learning:

The equipment helped.
I used it (the equipment) when I was stuck but not always.
I like using equipment.
It would be worse without equipment.
I used it (the equipment) when I was stuck but not always.

4.22 Participants' perceptions of the role of the equipment in learning

Section 4.20 suggests that participants did experience greater retrieval of facts after the intervention. This study is interested in the role of the equipment in learning these facts. The data from the semi-structured interviews show a range of answers. The two interviewees with the highest progress and achievement mentioned the visual aspect of using equipment:

I used Numicon because I can see what the number is.
I picture the pieces in my head.

One of the lower achieving interviewees discussed the importance of understanding and explanation during the intervention. Conceptual understanding was a key aspect of the intervention, and explanations were always accompanied with a visual model using Numicon and Cuisenaire rods:

The explanations helped me.

4.23 The teaching of additive understanding conceptually led to wider understanding of other areas of maths

Lucy and Layla both commented on the new learning that they acquired during this intervention. It is possible that the teaching of conceptual understanding helps learners to identify the interconnected nature of mathematics:

> We learnt extra stuff. We learnt about algebra.
> I learnt things I didn't know how to do like negative numbers 6 - 9.

4.24 The impact of the intervention on questions testing additive understanding

One of the limitations of the diagnostic test (Section 4.16) was that the researcher was not able to probe understanding or lack of understanding. The diagnostic questions were also quite different to the teaching intervention where the use of talk and explanations alongside the concrete models was routine. As part of the semi-structured interview, participants were given the calculation 2 - 6 = 8 and asked to check whether this was correct. The analysis of this question provides insight into the role of structured manipulatives when teaching additive reasoning. Layla (who had the highest performance on the fluency tests) was unsure about how to respond to 2 - 6 = 8:

> *Wrong? I don't know.*
>
> Evie who had the lowest performance provided this response:
>
> *It's wrong because it's not higher than 8. (Finds 8 and 6 and benchmarks the combined value of these against 10. Explains 2 should be 14).*

Section 4.4 discussed the variability in the sample; however, from Evie's and Layla's responses, it is important to consider whether individual understanding across areas of additive reasoning vary considerably. Lucy had middling achievement and responded with some ambiguity. The researcher then prompted her to use the equipment to clarify her thinking:

> *It's not right. They didn't use the equipment right. I don't know why (it is wrong).*

[Prompted to use the equipment]

You can't do it. There's only 2. (Picks up 6 and 8 to make 14). They did 8 - 6 = 2.

Both Lucy's and Evie's responses show how they are able to use the equipment to help justify and explain their thinking. Aiden's response showed that his own use of manipulatives provided feedback while replicating the principles of additive reasoning:

It's right.

[Prompted to check with equipment]

It's wrong.

The verbal diagnostic question provides additional data augmenting the results from Section 4.9 where participants had to find a missing minuend when provided with one of the related addition facts. During this phase of the data collection, only 20% of participants provided a correct answer with a mathematically correct explanation. Participants were much more successful identifying an incorrect minuend in the semi-structured interviews. There are several possible reasons for this. First, two of the participants were able to justify their thinking after being prompted by the interviewer. As part of these prompts, participants were encouraged to use the manipulatives. This helped Lucy to understand her mistake and find the correct answer. Using the equipment also helped Aiden to realise that the equation was not correct. Evie used the equipment without being prompted.

4.25 Limitations

A major limitation of the semi-structured interviews was the position of the interviewer as a teacher at the school of the study. Despite assurances otherwise, participants may have felt that certain answers were not permitted and therefore moderated or changed their answers. A second limitation is that the interviewer was the teacher carrying out the intervention. Participants may have consciously or unconsciously provided answers based on the actions or words of the teacher during the intervention.

Chapter 5

5.1 The impact of structured manipulatives on fluency

Research question 1 explored whether the use of structured manipulatives improved fluency of addition and subtraction facts. The current study identified a significant mean increase in participants' fluency scores. Due to the lack of a control group, it is difficult to ascertain the specific role of the manipulatives. It is possible that practice during the intervention resulted in improvement in fluency as previous studies have found this (Hopkins & Egeberg, 2009). Evidence from the semi-structured interviews suggests that the structured manipulatives may have had a role in improving fluency as participants spoke positively about the equipment. Another possibility is that the intervention improved understanding of additive reasoning, and consequently this led to an improvement in fluency. A considerable amount of research has examined the connected nature of additive reasoning and fluency (Section 2.11). The schema-based view promotes a model of LTM where additive reasoning knowledge affords a more organised storage of facts (Canobi et al., 1998). Future research might examine the relationship between structured manipulatives, fluency, and additive reasoning using a study with three interventions (manipulatives on fluency; manipulatives on additive reasoning; manipulatives on additive reasoning and fluency) to examine these relationships more closely.

Individual differences in fluency gains were observed and future research might explore if this type of intervention is more effective for some students than others. Section 4.4 highlights that the sample had considerable variance. The current study had broad knowledge of participants' mathematical achievement before sampling. Future research might select a sample using diagnostic testing. Another important area for future research would be to examine whether the improvement in fluency is maintained after the intervention phase, and whether any improvement then impacts on wider mathematical achievement.

5.2 The impact of structured manipulatives on additive reasoning

Research question 2 focused on whether additive reasoning improved with the use of structured manipulatives. The findings for this question were mixed. A significant improvement in mean correct answers was observed on the empty box test; however, the participants performed poorly on this test in comparison to the fluency test. This result supports previous research suggesting that these types of questions are challenging in this population (Crooks, 2010). The lack of improvement in the COF test was not expected. Future research might explicitly direct participants to the fact that questions were relationally ordered. Baroody (1999) states that this type of data collection method is dissimilar to learners' everyday experience of school where calculations generally stand alone and require solving in isolation. It is possible that the poorer performance on the COF test was due to participants using a different strategy without automaticity. A limitation of the current study is the lack of a self-report tool at the end of COF tests.

The significant mean improvement in the empty box tests supports the notion that additive reasoning improved in the sample. Some support for this comes from analysis of the diagnostic questions. Some improvements in areas of additive reasoning were identified; however, individual differences in participants' profiles of additive reasoning were also identified (Dowker, 2005). Furthermore, this research supports the body of literature suggesting that sub-concepts of additive reasoning vary in their difficulty. Commutativity was the sub-concept most easily acquired by learners, supporting previous research (Baroody et al., 1983; Canobi et al., 1998). It might be considered that commutativity is modelled effectively using structured manipulatives as the two parts can be physically reordered while maintaining the same total size (whole). Despite an improvement in understanding commutativity, it is possible that this principle was misapplied to other principles. Very few participants identified that subtraction was not commutative at post-test. The verbal diagnostic question asked during the semi-structured interview ($2 - 8 = 6$) examined participants' understanding of the importance of order during subtraction. Availability of the manipulatives appeared to allow participants access to, or to test, some of the conceptual understanding needed to answer this question. Future research might examine the results of diagnostic tests with and without manipulatives. It is possible that some learners had not fully internalised part-whole relationships, which would allow them to work abstractly with these principles, and required the manipulatives to test or check answers, providing feedback on their mathematical thinking.

It appears that participants misapplied commutativity when asked if an addition question helped solve a subtraction question with the same numbers. Research has highlighted concerns over the use of shortcuts compared to the use of principles (Canobi, 2005). Symbolically, commutativity is the subtle change in an equation ($a + b$, $b + a$). In the current study, participants were unlikely to identify that a similar subtle change in equations broke additive principles (a + b, a – b). The researcher noted some participants used quite basic, unspecific language to describe commutativity (*You just swap them around).* Alongside physical manipulation, future research might explore the use of explicit teaching of specific mathematically accurate definitions of additive principles.

Overall, participants appear to have found it challenging to understand the inverse relationship between addition and subtraction. Very few mathematically correct answers were present at post-test in the two diagnostic questions that required participants to transfer understanding from one operation to another using knowledge of the inverse relationships. Previous research has identified that this is a particularly challenging area of additive reasoning (Section 2.1). As the intervention was relatively short, it might be expected that this aspect of additive reasoning would not develop to the same degree as other areas. Research has shown greater effectiveness with manipulatives if they are used for a year (Sowell, 1989). Future research might examine whether a longer intervention improved participants' understanding of these inverse relationships.

Findings from the semi-structured interviews identified a preference for the use of Numicon over Cuisenaire rods because the visual value of Numicon was readily available to learners. Although it is not immediately obvious, the obscured value of Cuisenaire rods may be an affordance as part-whole structure, and regularity is not detracted from by the numerical value of the manipulative. Perceptual richness of manipulatives has been shown to detract from mathematical learning (Carbonneau, Wong, & Borysenko, 2020). It is possible that explicit teaching with Cuisenaire rods would have helped learners to identify additive reasoning structure more clearly.

Despite the majority of participants not being able to abstractly understand additive and subtraction inverse, there is some evidence that knowledge of the two operations and their relationships developed. No participants provided the related addition and two subtraction calculations to a given addition calculation at pre-test. At post-test, this increased to 60%. This is an important finding as the 60% of participants identified three correct calculations without misapplying rules or shortcuts. Furthermore, 20% of participants provided some correct calculations. In summary, some improvements in additive reasoning were identified, both using quantitative methods and through diagnostic questions. Future research might explore whether a longer intervention is required to impact on the more challenging aspects of additive reasoning. Research might also explore whether retrieval of facts from LTM needs to reach a level of mastery, freeing the working memory load to focus on additive reasoning (Baroody, 1999).

5.3 Participants' perceptions of the manipulatives and the intervention

Research questions 3 and 4 explore participants' perceptions of the manipulatives, and whether learning gains altered perceptions. Semi-structured interviews with selected participants used maximum variation sampling, and results show positive attitudes towards the intervention and the manipulatives, especially the Numicon shapes. This finding is important as participants of this age may have identified using equipment as being child-like. Furthermore, all participants interviewed had positive attitudes despite variation in their relative gains. It is possible that the low-achieving participants could identify improvements in their learning and understanding, but the data collection tools were not sensitive enough to capture those improvements. As participants felt positively about the manipulatives and their impact on learning, it is important to explore the affordances further. Participants appeared to benefit from the visual organisation of dots in the Numicon shape. The visual affordance of the Numicon shape may provide support to research highlighting the importance of subitising and developing part-whole relationships as a supplementary pedagogy for developing arithmetic alongside counting (Kullberg & Björklund, 2019; Young-Loveridge, 2011b). The use of manipulatives provides support for the modality effect aspect of CLT as the manipulatives may have decreased working memory load, allowing for greater recognition of part-whole relations. Future research might explore whether structured manipulatives further decrease cognitive load by encouraging participants not to count to calculate (de Chambrier et al., 2018).

Research question 5 explored whether counting strategies are frequently used by participants in this sample. The data from self-reports suggest that counting strategies were frequently used. However, it appears that there was a slight reduction in the use of counting strategies; a limitation of the present study is that this cannot not be quantified or explored further. During semi-structured interviews, all participants commented that they were counting less and had now had greater knowledge of addition and subtraction facts. An interesting finding was that participants who continued to self-report counting strategies at post-test had the lowest rates of fluency in the sample. This provides some evidence to suggest that participants who count to calculate are disadvantaged and counting may increase cognitive load when learning mathematics.

In summary, participants responded positively about the manipulatives and the intervention. The current study reported an improvement in fluency and additive reasoning from pre-test to post-test; however, within the sample, individual differences were evident.

Chapter 6

In conclusion, the present study aimed to explore whether the use of structured manipulatives improved fluency and additive reasoning in a group of lower-achieving year 7 and 8 students. There was evidence (Section 4.6) that counting strategies were prevalent in this sample at pre-test. Through the lens of CLT, the use of counting strategies may overload working memory, impeding the ability to identify additive relationships. The reliance on counting strategies may have a circular effect as consequently, participants may continue to count to calculate, resulting in the principles of additive reasoning not being discovered, and therefore, not being available to derive new addition and subtraction facts (Gray & Tall, 1994). In line with CLT, the researcher conjectured (Section 1.7) that structured manipulatives would alleviate the load on working memory during instructional time, providing participants with the attentional capacity to identify additive reasoning principles. This assertion builds on the work of Canobi (2005): “It seems likely that knowledge of addition principles emerges through noticing regularities in the ways in which physical objects can be combined” (p. 224). Linked to the wider literature (Section 2.3), the structured manipulative accurately represents the additive reasoning principles required to develop part-whole understanding. CLT also framed the study in a broader context as the researcher conjectured that use of counting strategies in mathematics lessons may lead to working memory overload.

Participants responded positively about the manipulatives and reported greater use of retrieval strategies post-intervention. Participants had an improved understanding of additive reasoning after taking part in the intervention.

6.1 Practical implications of the current study

Practical implications may be drawn from this research. A lack of fluency and additive reasoning was evident in the sample at pre-test. The majority of participants used counting strategies at the start of the intervention. Within the classroom, early identification of learners not displaying additive reasoning understanding and also relying on counting strategies may provide the teacher with the time and opportunity to develop more sophisticated strategies and understanding with these students. The data collection tools used in this study could be used as assessment tools to help teachers identify students lacking in additive reasoning and fluency. Such diagnosis is especially pertinent as the current study supports the literature which states that some aspects of additive reasoning are more difficult to acquire than others

(Section 5.2). Initial teacher education providers should ensure that additive reasoning and conceptual understanding are given adequate weight when training new practitioners. Moreover, as part of on-going formative assessment, educators should be cognizant of learners who become reliant on counting strategies rather than developing more sophisticated understanding. As well as identifying at risk learners with low fluency and incomplete additive reasoning understanding, educators also need tools to remedy this issue. Providing teachers with an understanding of CLT may highlight the importance of placing basic facts into LTM, and how structured manipulatives might scaffold learners to develop flexible but abstract understanding of additive principles. Educators and policymakers should consider the role of subitising and unitising alongside counting as pathways to more developed arithmetic. The current study suggests that structured manipulatives may have a role to play in the development of fluency and additive understanding. Educators should consider using structured manipulatives as a means of developing additive reasoning.

6.2 Recommendations for future research

It would be productive to examine whether extended exposure with structured manipulatives impacts on some of the more challenging areas of additive reasoning (Section 5.2). Further research would be helpful to examine whether the two types of structured manipulatives used in the current study vary in their efficacy. Also, it is recommended that further research examines the effectiveness of the structured manipulatives across larger samples.

A surprising finding from the current study was that students with the smallest relative improvements spoke positively about the manipulatives. The mixed-methods approach provided such insights in the current study that may have been obscured with a solely quantitative approach. The area of additive reasoning is akin to conceptual understanding. Therefore, in further studies, data collection using verbal diagnostic questions may provide even richer data into the understanding of learners and the development of additive reasoning.

A large body of research examines the bi-directional relationship of conceptual understanding and procedural understanding (Section 2.12). The present study did not measure whether gains in fluency led to improved additive reasoning or vice versa. Useful in future research therefore would be to examine the bi-directional nature of these two intertwined areas of mathematics to enable teachers and policy makers to further identify effective ways of teaching these key mathematical concepts. Finally, future research is needed to continue to broaden our understanding of the use of manipulative type in the learning of key mathematical concepts.

In conclusion, the present study has demonstrated how theory and practice can be blended to effect positive change for students at a crucial stage of their education. The intervention was relatively short in terms of session length and duration and the required materials were inexpensive. Such an intervention is therefore easily replicated by classroom teachers or teacher aides. An understanding of CLT drove the researcher to consider how working memory load can be reduced in the short-term (using structured manipulatives) and in the long-term (placing facts into LTM and illuminating the structure of additive relationships). The intervention and use of structured manipulatives provided learners with a deeper understanding of additive reasoning, and improved addition and subtraction fluency. These key aspects of arithmetic are potentially powerful for enhancing learners' future lives and life opportunities.

References

Adams, J. W., & Hitch, G. J. (1997). Working memory and children's mental addition. *Journal of Experimental Child Psychology, 67*(1), 21. doi:10.1006/jecp.1997.2397

Anthony, G. (2007). *Effective pedagogy in mathematics/pāngarau: Best evidence synthesis iteration (BES)*. Wellington, NZ: Ministry of Education.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44*(1-2), 75–106. doi:10.1016/0010-0277(92)90051-I

Baddeley, A. D. (2018). *Exploring working memory: Selected works of Alan Baddeley*. Abingdon, Oxon: Routledge.

Baddeley, A., & Hitch, G. (1974). A proposed working memory system. *The Psychology of Learning and Motivation, 8*, 74.

Baroody, A. J. (1987). The development of counting strategies for single-digit addition. *Journal for Research in Mathematics Education, 18*(2), 141–157. doi:10.2307/749248

Baroody, A. J. (1994). An evaluation of evidence supporting fact-retrieval models. *Learning and Individual Differences, 6*(1), 1–36. doi:10.1016/1041-6080(94)90013-2

Baroody, A. J. (1999). Children's relational knowledge of addition and subtraction. *Cognition and Instruction, 17*(2), 137–175. doi:10.1207/S1532690XCI170201

Baroody, A. J., & Dowker, A. (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise*. Mahwah, NJ: Lawrence Erlbaum.

Baroody, A. J., Ginsburg, H. P., & Waxman, B. (1983). Children's use of mathematical structure. *Journal for Research in Mathematics Education, 14*(3), 156–168. doi:10.2307/748379

Baumann, C., & Winzar, H. (2016). The role of secondary education in explaining competitiveness. *Asia Pacific Journal of Education, 36*(1), 13–30. doi:10.1080/02188791.2014.924387

Bonne, L. (2003). *Numeracy equipment and year 3 children: "Bright, shiny stuff" or supporting the development of part-whole thinking?* (Master’s thesis). Wellington, NZ:Wellington College of Education.

Bouck, E. C., & Park, J. (2018). A systematic review of the literature on mathematics manipulatives to support students with disabilities. *Education & Treatment of Children, 41*(1), 65. doi:10.1353/etc.2018.0003

Brownell, W. (1945). Teaching of mathematics in grades I thru VI. *Review of Educational Research, 15*(1), 276.

Burns, M. K., Codding, R. S., Boice, C. H., & Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review, 39*(1), 69.

Bynner, J., & Parsons, S. (2006). *Does numeracy matter more?* National Research and Development Centre for adult literacy and numeracy, Institute of Education, University of London.

Canobi, K. H. (2005). Children's profiles of addition and subtraction understanding. *Journal of Experimental Child Psychology, 92*(3), 220–246. doi:10.1016/j.jecp.2005.06.001

Canobi, K. H. (2009). Concept–procedure interactions in children's addition and subtraction. *Journal of Experimental Child Psychology, 102*(2), 131–149. doi:10.1016/j.jecp.2008.07.008

Canobi, K. H., Reeve, R. A., & Pattison, P. E. (1998). The role of conceptual understanding in children's addition problem solving. *Developmental Psychology, 34*(5), 882–891.

Canobi, K. H., Reeve, R. A., & Pattison, P. E. (2002). Young children's understanding of addition concepts. *Educational Psychology, 22*(5), 513–532. doi:10.1080/0144341022000023608

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology, 105*(2), 380–400. doi:10.1037/a0031084

Carbonneau, K. J., Wong, R. M., & Borysenko, N. (2020). The influence of perceptually rich manipulatives and collaboration on mathematic problem-solving and perseverance. *Contemporary Educational Psychology, 61*. doi:10.1016/j.cedpsych.2020.101846

Carpenter, T., Hiebert, J., & Moser, J. (1983). The effect of instruction on children's solutions of addition and subtraction word problems. *Educational Studies in Mathematics, 14*(1), 55–72. doi:10.1007/BF00704702

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education, 15*(3), 179–202. doi:10.2307/748348

Chamberlain, M., & Caygill, R. (2012). *Key findings from New Zealand's participation in the Progress in International Reading Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS) in 2010/11*. Wellington, NZ: Research Division, Ministry of Education.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332. doi:10.1207/s1532690xci0804_2

Cheng, Z.-J. (2012). Teaching young children decomposition strategies to solve addition problems: An experimental study. *Journal of Mathematical Behavior, 31*(1), 29–47. doi:10.1016/j.jmathb.2011.09.002

Ching, B. H.-H., & Nunes, T. (2017a). Children's understanding of the commutativity and complement principles: A latent profile analysis. *Learning and Instruction, 47*, 65–79. doi:10.1016/j.learninstruc.2016.10.008

Ching, B. H.-H., & Nunes, T. (2017b). The importance of additive reasoning in children's mathematical achievement: A longitudinal study. *Journal of Educational Psychology, 109*(4), 477–508. doi:10.1037/edu0000154

Ching, B. H.-H., & Wu, X. (2019). Concreteness fading fosters children's understanding of the inversion concept in addition and subtraction. *Learning and Instruction, 61*, 148–159. doi:10.1016/j.learninstruc.2018.10.006

Cho, S., Ryali, S., Geary, D. C., & Menon, V. (2011). How does a child solve 7 + 8? Decoding brain activity patterns associated with counting and retrieval strategies. *Developmental Science, 14*(5), 989–1001. doi:10.1111/j.1467-7687.2011.01055.x

Clements, D. H., Sarama, J., & MacDonald, B. L. (2019). Subitizing: The neglected quantifier. In A. Norton & M. W. Alibali (Eds.), *Constructing number: Merging perspectives from psychology and mathematics education* (pp. 13–45). Cham: Springer International.

Cowan, R., Donlan, C., Shepherd, D. L., Cole-Fletcher, R., Saxton, M., & Hurry, J. (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *Journal of Educational Psychology, 103*(4), 786–803. doi:10.1037/a0024556

Creswell, J. W. (2011). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Los Angeles: Sage.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks: Sage.

Crooks, T. (2010). *Mathematics: Assessment results 2009*. Dunedin, NZ: Educational Assessment Research Unit, University of Otago, under contract to the Ministry of Education.

de Chambrier, A.-F., Thevenot, C., Barrouillet, P., & Zesiger, P. (2018). Frequency of finger looking during finger counting is related to children's working memory capacities. *Journal of Cognitive Psychology, 30*(5-6), 503–510. doi:10.1080/20445911.2018.1502190

Department for Education. (2013). *National curriculum*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/335158/PRIMARY_national_curriculum_-_Mathematics_220714.pdf.

Ding, M., & Auxter, A. (2017). Children's strategies to solving additive inverse problems: A preliminary analysis. *Mathematics Education Research Journal, 29*(1), 73–92. doi:10.1007/s13394-017-0188-4

Dowker, A. (2005). *Individual differences in arithmetic implications for psychology, neuroscience, and education*. New York, NY: Psychology Press.

Dowker, A. (2014). Young children's use of derived fact strategies in addition and subtraction. *Frontiers in Human Neuroscience, 7,* 924. doi:10.3389/fnhum.2013.00924

Drabble, S., & O'Cathain, A. (2015). In S. N. Hesse-Biber & B. Johnson (Eds.). *The Oxford handbook of multimethod and mixed methods research inquiry* Oxford: Oxford University Press.

Faulkner, V., & Ainslie, J. (2017). Subitising through the years.(Essay). *Australian Primary Mathematics Classroom, 22*(1), 28.

Findell, B., Kilpatrick, J., & Swafford, J. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Forder, K. (2016). *A study to evaluate the impact of a Numicon-based intervention on the numeracy attainment and attitude towards numeracy of children in lower key stage 2 identified as experiencing difficulties in mathematics*. ProQuest Dissertations.

Ganley, C. M., & Vasilyeva, M. (2014). The role of anxiety and working memory in gender differences in mathematics. *Journal of Educational Psychology*. *106*(1), 105.

Geary, D. C. (1994). *Children's mathematical development research and practical applications* (1st ed.). Washington, DC: American Psychological Association.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & Catherine Desoto, M. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology, 88*(2), 121–151. doi:10.1016/j.jecp.2004.03.002

Gilmore, C., Clayton, S., Cragg, L., McKeaveney, C., Simms, V., & Johnson, S. (2018). Understanding arithmetic concepts: The role of domain-specific and domain-general skills. *PLoS ONE, 13*(9), e0201724. doi:10.1371/journal.pone.0201724

Gilmore, C. K., & Spelke, E. S. (2008). Children's understanding of the relationship between addition and subtraction. *Cognition, 107*(3), 932–945. doi:10.1016/j.cognition.2007.12.007

Gray, E. (1991). An analysis of diverging approaches to simple arithmetic: Preference and its consequences. *Educational Studies in Mathematics, 22*(6), 551–574. doi:10.1007/BF00312715

Gray, E. M., & Tall, D. O. (1994). Duality, ambiguity, and flexibility: A "proceptual" view of simple arithmetic. *Journal for Research in Mathematics Education, 25*(2), 116–140. doi:10.2307/749505

Green, M., Piel, J. A., & Flowers, C. (2008). Reversing education majors' arithmetic misconceptions with short-term instruction using manipulatives. *The Journal of Educational Research, 101*(4), 234-242. doi:10.3200/JOER.101.4.234–242

Hahn, R. A., & Chattopadhyay, S. K. (2019). Linking studies to assess the life expectancy associated with eighth grade school achievement. *Preventive Medicine Reports, 16*. doi:10.1016/j.pmedr.2019.100980

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. doi:10.3102/003465430298487

Haylock, D. (2010). *Mathematics explained for primary teachers* (4th ed.). Los Angeles: Sage.

Higgins, J., & Parsons, R. (2009). A successful professional development model in mathematics: A system-wide New Zealand case. *Journal of Teacher Education, 60*(3), 231–242. doi:10.1177/0022487109336894

Hopkins, S., & Egeberg, H. (2009). Retrieval of simple addition facts: Complexities involved in addressing a commonly identified mathematical learning difficulty. *Journal of Learning Disabilities, 42*(3), 215–229. doi:10.1177/0022219408331041

Jesica, F., Silvia, J., Irene, I.-R., & Juan Pablo, B. (2018). Resolution of arithmetic problems, processing speed and working memory in children. *Trends in Psychology, 26*(3), 1249–1266. doi:10.9788/tp2018.3-05en

Johnson, B., & Christensen, L. (2014). *Educational research: Quantitative, qualitative, and mixed approaches* (5th ed.). Thousand Oaks, CA: Sage.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology, 85*(2), 103–119. doi:10.1016/S0022-0965(03)00032-8

Koponen, T. K., Sorvo, R., Dowker, A., Räikkönen, E., Viholainen, H., Aro, M., & Aro, T. (2018). Does multi-component strategy training improve calculation fluency among poor performing elementary school children? *Frontiers in Psychology, 9*, 1187, 1-14. doi:10.3389/fpsyg.2018.01187

Kullberg, A., & Björklund, C. (2019). Preschoolers' different ways of structuring part-part-whole relations with finger patterns when solving an arithmetic task. *Zdm - The International Journal on Mathematics Education*. doi:10.1007/s11858-019-01119-8

Kurumeh, M. (2010). Effect of cuisenaire rods' approach on students' interest in decimal fractions in junior secondary school: Makurdi Metropolis. *Global Journal of Educational Research, 9*(1/2), 19–25. doi:10.4314/gjedr.v9i1-2.62517

Lee, J. (2010). Tripartite growth trajectories of reading and math achievement: Tracking national academic progress at primary, middle, and high school levels. *American Educational Research Journal, 47*(4), 800–832. doi:10.3102/0002831210365009

Manches, A., & O'Malley, C. (2016). The effects of physical manipulatives on children's numerical strategies. *Cognition and Instruction, 34*(1), 27–50. doi:10.1080/07370008.2015.1124882

McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74*(3), 240–260. doi:10.1006/jecp.1999.2516

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97. doi:10.1037/h0043158

Ministry of Education, NZ (2008). *The number framework* (Rev. ed. 2007.). Wellington, NZ:Author.

Ministry of Education. (2012). *Mathematics syllabus primary one to five*. Retrieved from https://www.moe.gov.sg/docs/default-source/document/education/syllabuses/sciences/files/primary_mathematics_syllabus_pri1_to_pri5.pdf

Ministry of Education, NZ (2017). Mathematics achievement: What we know from New Zealand's participation in TIMSS 2014/15 and PISA 2015: Wellington, NZ: Author.

Ministry of Education, NZ (2018). *Achievement and progress in mathematics, reading and writing in primary schooling*. Retrieved from https://www.educationcounts.govt.nz/__data/assets/pdf_file/0019/185023/20171213-Achievement-and-Progress-in-mathematics-reading-and-writing.pdf.

Mulligan, J., & Mitchelmore, M. (2009). Awareness of pattern and structure in early mathematical development. *Mathematics Education Research Journal, 21*(2), 33–49. doi:10.1007/BF03217544

Nelson, P. M., Burns, M. K., Kanive, R., & Ysseldyke, J. E. (2013). Comparison of a math fact rehearsal and a mnemonic strategy approach for improving math fact fluency. *Journal of School Psychology, 51*(6), 659–667. doi:10.1016/j.jsp.2013.08.003

Newby, P. (2010). *Research methods for education*. Harlow, England: Pearson Education.

Nunes, T., Bryant, P., Barros, R., & Sylva, K. (2012). The relative importance of two different mathematical abilities to mathematical achievement. *The British Journal of Educational Psychology, 82*(1), 136–156. doi:10.1111/j.2044-8279.2011.02033.x

Nunes, T., Bryant, P., Hallett, D., Bell, D., & Evans, D. (2009). Teaching children about the inverse relation between addition and subtraction. *Mathematical Thinking and Learning, 11*(1-2), 61–78. doi:10.1080/10986060802583980

Ostad, S. A. (1997). Developmental differences in addition strategies: A comparison of mathematically disabled and mathematically normal children. *British Journal of Educational Psychology, 67*(3), 345–357. doi:10.1111/j.2044-8279.1997.tb01249.x

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Prather, R. W., & Alibali, M. W. (2009). The development of arithmetic principle knowledge: How do we know what learners know? *Developmental Review, 29*(4), 221–248. doi:10.1016/j.dr.2009.09.001

Resnick, L. B. (1984). *A developmental theory of number understanding*. Washington, DC: Distributed by ERIC Clearinghouse.

Resnick, L. B. (1992). From protoquantities to operators: Building mathematical competence on a foundation of everyday knowledge. In G. Leinhardt, R. Putnam, & R. A. Hattrup (Eds.), *Analysis of arithmetic for mathematics teaching* (pp. 373–429). Hillsdale, NJ: Lawrence Erlbaum.

Rittle-Johnson, B., Siegler, R. S., & Wagner Alibali, M. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*(2), 346–362. doi:10.1037/0022-0663.93.2.346

Sammons, P. (2010). The contribution of mixed methods to recent research on educational effectiveness. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (2nd ed., pp. 697-723). Thousand Oaks, CA: Sage.

Schneider, W., & Niklas, F. (2017). Intelligence and verbal short-term memory/working memory: Their interrelationships from childhood to young adulthood and their impact on academic achievement. *Journal of Intelligence, 5*(2), 1-19. doi:10.3390/jintelligence5020026

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4. doi:10.3102/0013189X015002004

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116*(3), 250-264. doi:10.1037/0096–3445.116.3.250

Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education, 20*(5), 498–505. doi:10.2307/749423

Starkey, G. S., & McCandliss, B. D. (2014). The emergence of "groupitizing" in children's numerical cognition. *Journal of Experimental Child Psychology, 126*, 120–137. doi:10.1016/j.jecp.2014.03.006

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory* (Vol. 1). New York, NY: Springer New York.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185–233. doi:10.1207/s1532690xci1203_1

Thompson, P. W. (1994). Concrete materials and teaching for mathematical understanding. *The Arithmetic Teacher, 41*(9), 556.

Torbeyns, J., Peters, G., De Smedt, B., Ghesquière, P., & Verschaffel, L. (2016). Children's understanding of the addition/subtraction complement principle. *British Journal of Educational Psychology, 86*(3), 382–396. doi:10.1111/bjep.12113

Van Den Hout, M. A., Engelhard, I. M., Smeets, M. A. M., Hornsveld, H., Hoogeveen, E., De Heer, E., . . . Rijkeboer, M. (2010). Counting during recall: Taxing of working memory and reduced vividness and emotionality of negative memories. *Applied Cognitive Psychology, 24*(3), 303–311. doi:10.1002/acp.1677

Walker, D., Bajic, D., Mickes, L., Kwak, J., & Rickard, T. C. (2014). Specificity of children's arithmetic learning. *Journal of Experimental Child Psychology, 122*(1), 62–74. doi:10.1016/j.jecp.2013.11.018

Wing, T., & Tacon, R. (2007). Teaching number skills and concepts with Numicon materials. *Down's Syndrome, Research and Practice: The Journal of the Sarah Duffen Centre, University of Portsmouth, 12*(1), 22–26. doi:10.3104/practice.2018

Wubbena, Z. C. (2013). Mathematical fluency as a function of conservation ability in young children. *Learning and Individual Differences, 26*, 153.

Young-Loveridge, J. M. (1999). The acquisition of numeracy. *Set: Research Information for Teachers,* 1, 1-8.

Young-Loveridge, J. (2002). Early childhood numeracy: Building an understanding of part-whole relationships. *Australian Journal of Early Childhood, 27*(4), 36–42.

Young-Loveridge, J. (2011a). Assessing the mathematical thinking of young children in New Zealand: The initial school years. *Early Child Development and Care, 181*(2), 267–276. doi:10.1080/03004430.2011.536645

Young-Loveridge, J. (2011b). Rethinking the role of counting in mathematics learning. *Teachers and Curriculum, 12*(1), 79. doi:10.15663/tandc.v12i1.34

Young-Loveridge, J., & Bicknell, B. (2015). Number-fact knowledge and mathematical problem-solving of 5- to 7-year-olds. *Asia-Pacfic Journal of Research in Early Childhood Education, 9*, 27–43. doi:10.17206/apjrece.2015.9.3.27

Zamarian, L., Ischebeck, A., & Delazer, M. (2009). Neuroscience of learning arithmetic: Evidence from brain imaging studies. *Neuroscience and Biobehavioral Reviews, 33*(6), 909–925. doi:10.1016/j.neubiorev.2009.03.005

Appendices

Appendix A

Fluency pre-test/post-test

Please answer these questions.

Try them all in order by going down the columns. Skip any questions you cannot do

It does not matter if you don't know the answers.

7 + 8 =	7 + 5 =	13 – 8 =
15 – 9 =	18 – 9 =	6 + 7 =
6 + 9 =	9 + 3 =	16 – 9 =
17 – 9 =	17 – 8 =	9 + 4 =
6 + 5 =	8 + 9 =	12 – 8 =
11 – 8 =	14 – 5 =	7 + 5 =
6 + 7 =	4 + 8 =	14 – 7 =
13 – 6 =	15 – 7 =	6 + 4 =
8 + 4 =	6 + 9 =	13 – 4 =
15 – 6 =	12 – 7 =	5 + 9 =
4 + 9 =	8 + 3 =	16 – 8 =
11 – 6 =	4 + 7 =	4 + 9 =
4 + 7 =	13 – 9 =	14 – 8 =
18 – 9 =	5 + 9 =	15 – 9 =
9 + 6 =	12 – 8 =	9 + 3 =
12 – 9 =	9 + 2 =	16 – 7 =
8 + 5 =	14 —6 =	7 + 3 =
13 – 7 =	6 + 5 =	14 – 9 =
5 + 9 =	7 + 5 =	17 – 6 =
11 – 2 =	12 – 6 =	13 – 5 =

How did you solve these?

Appendix B

Empty box pre-test/post-test

$\boxed{} - 8 = 5$	$\boxed{} - 7 = 9$	
$5 + \boxed{} = 16$	$6 + \boxed{} = 15$	
$12 - \boxed{} = 4$	$15 - \boxed{} = 6$	
$\boxed{} + 9 = 16$	$\boxed{} + 4 = 12$	
$\boxed{} - 9 = 7$	$\boxed{} - 8 = 5$	
$7 + \boxed{} = 16$	$6 + \boxed{} = 13$	
$14 - \boxed{} = 5$	$14 - \boxed{} = 9$	
$\boxed{} + 7 = 11$	$\boxed{} + 9 = 15$	
$\boxed{} - 4 = 9$	$\boxed{} - 6 = 7$	
$4 + \boxed{} = 13$		
$11 - \boxed{} = 5$		
$\boxed{} + 6 = 15$		
$\boxed{} - 8 = 6$		
$9 + \boxed{} = 14$		
$12 - \boxed{} = 4$		
$\boxed{} + 7 = 14$		
$\boxed{} - 6 = 6$		
$9 + \boxed{} = 11$		
$13 - \boxed{} = 8$		
$\boxed{} + 5 = 13$		

Appendix C

Conceptually ordered fluency test

Please answer these questions.

Try them all in order by going down the columns. Skip any questions you cannot do

It does not matter if you don't know the answers.

6 + 9 =	12 — 9 =	13 – 8 =
9 + 6 =	9 + 3 =	5 + 8 =
15 — 9 =	12 – 3 =	13 – 5 =
15 – 6 =	3 + 9 =	8 + 5 =
17 – 9 =	7 + 8 =	12 – 8 =
17 – 8 =	15 – 8 =	11 — 7 =
9 + 8 =	8 + 7 =	11 – 4 =
8 + 9 =	15 – 7 =	7 + 4 =
8 + 4 =	5 + 6 =	4 + 7 =
12 – 4 =	11 — 6 =	7 + 9 =
4 + 8 =	6 + 5 =	16 – 7 =
12 – 8 =	11— 5 =	9 + 7 =
14 — 8 =	13 – 9 =	11 – 8 =
6 + 8 =	4 + 9 =	11 – 3 =
14 – 6 =	13 — 4 =	8 + 3 =
8 + 6 =	9 + 4 =	3 + 8 =
16 – 9 =	14 — 9 =	7 + 3 =
9 + 7 =	9 + 5 =	12 – 8 =
16 – 7 =	14 — 5 =	12 – 3 =
7 + 9 =	5 + 9 =	8 + 3 =

Appendix D

Diagnostic questions

If Mark knows 177 + 383 = 560 can he use this to solve 383 + 177?

Yes ☐ No ☐ Not sure ☐

Explain your answer

If Mark knows 177 + 383 = 560 can he use this to solve 383 – 177?

Yes ☐ No ☐ Not sure ☐

Explain your answer

If Mark knows the answer to 577 – 434, can he also solve 434 – 577.
Explain your answer

Yes ☐ No ☐ Not sure ☐

If Mark knows that 234 + 574 = 808, can he use this to solve the problem below?

$\square - 234 = 574$

Yes ☐ No ☐ Not sure ☐

Explain your answer

If Mark knows that 1 – 0.12 = 0.88, can he you use this to solve the problem below?

$0.12 + \square = 1$

Yes ☐ No ☐ Not sure ☐

Explain your answer

If Mark knows, 398 + 777 = 1175 what other number sentences can he make with these three numbers?

Appendix E

Interview protocol

Interview protocol

Introduction/Consent/Confidentiality

'Thank you for coming. The aim of this interview is to find how you feel about the mathematics intervention. I will make notes on what you say but your name will not be recorded. Everything you say is confidential. Also, there no right or wrong answers here. You will not get in trouble for anything you say. I will read your answers back to you to check you are happy with what I have written.
You can stop this interview at any point. Are you happy to continue?

1) How did you feel about the intervention?

Response A

2) Can you explain why you felt RESPONSE A about the intervention?

Response B

3) What effect do you think the intervention has had on your maths?

Response C

4) (Point to manipulatives) Can you talk to me about these?

Response D

Check all 4 responses

Thank student for responses. Assure students of confidentiality and anonymity

Appendix F

Sample

Will using equipment help understand maths?

INFORMATION SHEET FOR PARTICIPANTS

You are invited to take part in this research on improving maths. Please read this information before deciding whether or not you are happy to be in the study. If you decide to be in it, thank you. If you decide not to, thank you for thinking about this.

Who am I?

My name is Mr Daniel Green and I am a Masters student in Education at Victoria University of Wellington. This research project is work for my study. I also teach

What is the aim of the project?

This project about is about improving maths. The study involves learning maths facts using equipment.

I would like to test a strategy for teaching additive reasoning. This is specific part of maths focusing on addition and subtraction. I am planning to meet with a group of children from a range of classes for an hour after school for seven weeks. Your teacher has suggested that you might benefit from being involved in the research to evaluate the strategy. This is an invitation only and you do not have to participate.

This research has been approved by the Victoria University of Wellington Human Ethics Committee 0000026445.

How can you help?

You have been invited to help your maths learning. There are several parts to this study. First, you and the rest of the group will do a short test. This will be after school. This will help me to see if the project will help you. Some people who take the test will be part of the study. We will discuss whether you want to continue to take part. For the next seven weeks, you (and the rest of group) will work on maths with me for 30 minutes after school once a week. I will video record these sessions with your permission. The videos will help me to understand how well the project is helping your learning. After the seven weeks, you and the group will take another test (also after school). This will help me to check how well the project has helped your learning.

I will interview a few members of the group in Room 6, with their permission. I will ask questions about how you felt the study went. The interview will take no more than fifteen minutes and I will write up notes as you are talking. You can choose to not answer any question or stop the interview at any time, without giving a reason. You can withdraw from the study by contacting me before 20/11/18. Leaving the study early will not mean you are in trouble.

To thank you for participation, you will receive $20 voucher (koha) at the beginning of the study.

What will happen to the information you give?

This research is confidential*. This means that only me and my university supervisor will know who you are. Your information will be mixed with everyone else's and your name will not be used at all.

Only my supervisors and myself will read the notes of the interview and be able to watch the video footage . The interview transcripts, summaries and any recordings will be kept securely and destroyed on 30/06/2019

What will the project produce?

The information from my research will be used in a Masters thesis. Results may be published in academic or professional journals for teachers or discussed at academic or professional conferences.

If you accept this invitation, what are your rights as a research participant?

You do not have to accept this invitation if you don't want to. If you do decide to participate, you have the right to:

- choose not to answer any question;
- ask for the recorder to be turned off at any time during the maths session
- withdraw from the study before 20/11/18;
- ask any questions about the study at any time;
- receive a copy of your interview notes
- read over and comment on a written summary of your interview
- be able to read any reports of this research by emailing the researcher to request a copy.

If you have any questions or problems, who can you contact?

If you have any questions, either now or in the future, please feel free to contact [either/me]:

Student: Daniel Green	**Supervisor: Robin Averill**
University email address:	Role: Assistant Professor
	School: Education

* Confidentiality will be preserved except where you disclose something that causes me to be concerned about a risk of harm to yourself and/or others.