Calculation fluency: A mixed methods study in English Y6 primary classrooms

Emma-Louise Lord

Clare Hall

June 2019

This dissertation is submitted for the degree of Doctor of Philosophy

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Emma-Louise Lord

Title of study: Calculation fluency: A mixed methods study in English primary classrooms

The teaching and learning of written calculation strategies remains a high priority in many national curricula. However, the available literature was dominated by quantitative studies which explored a restricted range of arithmetic operations, paying limited attention to the role of confidence or the reasons behind the selection of their calculation strategies with the learners themselves. The literature revealed that calculation fluency was generally accepted to require flexibility, efficiency, accuracy, and conceptual understanding, yet recent curriculum reforms in English schools prioritised formal algorithms and thus appeared to restrict calculation fluency in the classroom. This study explored calculation fluency among the first cohort of Year 6 learners (10- to 11-year-olds) studying under the reforms by asking: *To what extent does calculation fluency among Year 6 learners vary by gender, confidence level and prior attainment?*

Phase 1 of this sequential mixed methods explanatory study involved a large-scale survey (N = 590) where each participant was presented with a ten-question, Likert-style mathematics confidence questionnaire followed by a workbook specifically developed for this study containing 16 age-related, context-free multi-digit calculations covering addition, subtraction, multiplication and division. The participants' answers were compared by gender, mathematical confidence and previous attainment using regression analysis. Phase 2 consisted of a purposeful sample of learners drawn from Phase 1 (n = 23) who attended individual semi-structured interviews exploring their workbook responses in more depth. Their comments were examined using framework analysis, then the findings from both phases were integrated together to address the research question.

The findings indicated that too many learners failed to satisfy the stated criteria for calculation fluency. Many learners worked inaccurately, inflexibly and inefficiently by prioritising formal algorithms irrespective of the merits of individual calculations. They either failed to recognise situations where other strategies might have been more efficient and less likely to lead to error, or they were unwilling to deviate from using formal algorithms. The findings indicated that confidence, rather than gender or prior attainment, had the greatest effect of the three predictor variables on use of the formal algorithm; confident learners were less likely to deviate from using formal algorithms than other learners. The findings also revealed that girls were significantly less likely to deviate from using formal algorithms than boys. However, most of the variance in calculation fluency was determined by factors other than gender, prior attainment or confidence. Calculation fluency was also affected by practice, knowledge of testing procedures and an individual's checking procedures.

Hence, it is recommended that future researchers consider adopting a mixed methods research design due to the insights gained in this thesis. Moreover, by addressing all four operations, I was able to identify patterns in my findings across the operations. Regarding policymakers, my findings indicated that the decision to prioritise formal algorithms in the primary curriculum may need further consideration. Schools should consider encouraging their learners to calculate more flexibly to increase their accuracy rates, calculation efficiency and conceptual understanding. Further research should be undertaken to ascertain the longer-term effects on both genders and differing mathematical confidence levels of limiting calculation flexibility at primary level when those learners will experience a curriculum dominated by problem-solving, rather than calculation, later in their education.

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed 80, 000 word limit for the Faculty of Education.

Acknowledgements

My research would have been impossible without the support of my family, friends, colleagues and the schools who participated in both the pilot project and subsequent study.

I would like to thank my supervisor, Dr Andreas Stylianides, for his guidance throughout this project. In particular, I would like to express my gratitude for his continual focus on developing future researchers by encouraging his students to contribute towards research groups and conferences throughout their studies. I would also like to thank Dr Ben Alcott for his support during the analysis stage of this study, especially for sharing effective ways to disseminate my findings for my intended audience.

This project would have been impossible without the teachers and schools who contributed towards the drafting process, the pilot study or the full study. Thank you all for your support, I hope that my findings will offer the insights that I hoped to gain when I planned this project.

I would also like to thank my colleagues, family and friends. When I joined NRICH, I had already embarked on my project. Nevertheless, the team has been unwavering in their support and ensured that I had the time and space to continue my studies. Most of all, I would like to thank my family and friends who have consistently supported this project even though it inevitably affected my availability for family get-togethers, days out and holidays. I hope that when you read this thesis you will not only begin to understand why I found this project so fascinating but also appreciate your key roles in making it happen. Thank you all.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACME	Advisory Committee on Mathematics Education
ASCM	Adaptive Strategy Choice Model
ATM	Association of Teachers of Mathematics
BEI	British Education Index
BERA	British Educational Research Association
CDQ	Confidence and Doubt Questionnaire
DCSF	Department for Children, Schools and Families
DfE	Department for Education
DfES	Department for Education and Science
DfEE	Department for Education and Employment
ERIC	Educational Resources Information Center
EThOS	E-theses Online Service
GCSE	General Certificate of Secondary Education
IRA	Inter-rater agreement
KS1	Key Stage One
KS2	Key Stage Two
L1	Level 1
L2	Level 2
L3	Level 3
L4	Level 4
L5	Level 5
LA	Local Authority
MA	Mathematical Association
MLR	Multiple Linear Regression
MaST	Mathematics Specialist Teacher
MLD	Mathematical Learning Disability
MMR	Mixed Methods Research
MP	Member of Parliament
NCETM	National Centre for Excellence in Teaching Mathematics
NCTM	National Council of Teachers of Mathematics
OECD	Organisation for Economic Co-operation
Ofsted	Office for Standards in Education, Children’s Services and Skills

OUP	Oxford University Press
PISA	Programme for International Student Assessment
QCA	Qualifications and Curriculum Authority
RQ	Research Question
SATs	Standard Assessment Tests
SIS	School Improvement Sevice
SLT	Senior Leadership Team
STA	Standards and Testing Agency
TIMSS	Trends in International Mathematics and Science Study
UIC	Unique Identification Code
Y2	Year 2
Y5	Year 5
Y6	Year 6
Y8	Year 8

Chapter One: Introduction

Introduction

Mathematics is an incredibly creative, flexible subject. It enables architects to design awe-inspiring buildings, scientists to save lives by predicting the spread of infectious diseases and cryptographers to protect our personal data and national security. A willingness to work flexibly enabled 18th century mathematician Leonhard Euler to resolve the Königsberg bridge problem concerning the possibility of traversing seven bridges in a city divided into four regions; realising that conventional geometry and algebra were insufficient for the challenge, his innovative solution drew upon Leibnitz's geometry of position. Euler's solution provided the foundations of graph theory and, much later, our modern internet. Over a hundred years after Euler, Florence Nightingale realised the potential of visual graphics for influencing public opinion during the Crimean War and her innovation transformed the science of statistics.

By moving seamlessly between different mathematical ideas, both Swiss-based mathematician Euler and English social reformer Nightingale were arguably demonstrating their mathematical fluency. In particular, their actions appeared to highlight the importance of flexibility. Nevertheless, the literature review presented in Chapter Two exposed inconsistent expectations in the UK and beyond regarding the relative importance of flexibility as well as other aspects of mathematical fluency. More specifically, the review will highlight the deep tensions between the advocates of traditional and reform curricula regarding the relative importance of procedural fluency and conceptual understanding as well as acknowledging more recent attempts to broach a compromise between the two seemingly irreconcilable camps. The review will also indicate that flexibility was not universally regarded as an essential aspect of working fluently in mathematics.

Recent reforms by the Department for Education (DfE, 2013a) appeared to reflect the compromises described in the above paragraph as well as the inconsistent expectations regarding the various aspects of mathematical fluency. Achieving conceptual understanding alongside procedural fluency were two of the stated aims of the reforms regarding number and calculation work, yet the new curriculum also restricted opportunities for primary-aged learners to work flexibly since their schools were compelled to prioritise formal algorithms (Figure 1) over and above any other calculation strategies. Following the introduction of the reforms (DfE, 2013a), corresponding adjustments were made to the assessment procedures for Year 6 (Y6) learners. The existing Standard Assessment Tests (SATs) were revamped; the mental mathematics paper was replaced by a written arithmetic paper. The accompanying mark scheme (Appendix A), issued by the Standards and Testing Agency (STA, 2016a, p. 13), favoured formal algorithms over and above all other strategies; incorrect calculations would be assessed for awarding half marks if and only if learners had attempted a formal algorithm. Over-emphasising the perceived importance of formal algorithms and number recall might mean that younger learners will not appreciate the value of treating each calculation on its individual merits rather than adopting a ‘one size fits all’ approach.

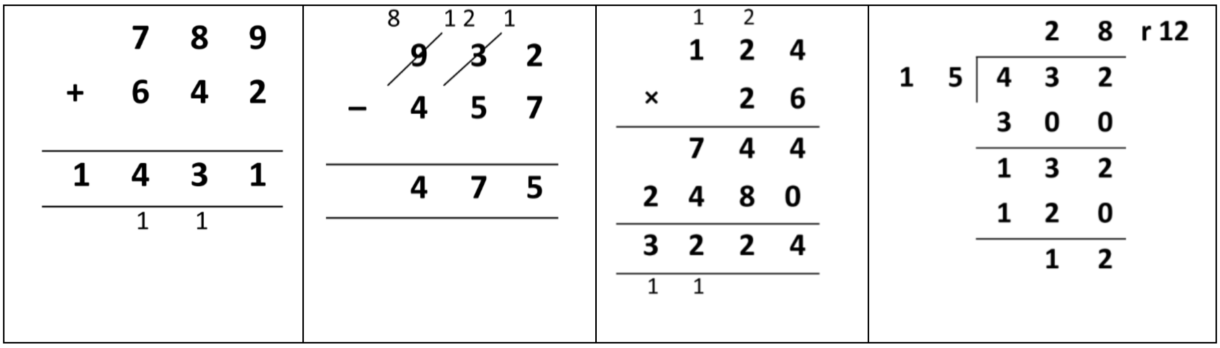


Figure 1. Example of a formal algorithm for each of the four operations. Adapted from *Mathematics programmes of study: Key stages one and two, national curriculum in England (pp. 46-47)* by DfE, 2013a.

Rationale

This mixed methods study explored calculation fluency with the first cohort of English Y6 learners studying under the reforms (DfE, 2013a). Those reforms appeared to have an almost immediate impact on classroom pedagogy: “Many schools are already interpreting this as practice, practice, practice of formal algorithms” (McClure, 2014, para. 3). The mathematical fluency of primary-aged learners studying under the reforms demanded further investigation. This thesis focused on their calculation fluency, a specific aspect of mathematical fluency which was explored in more depth in the literature review. This section of the introductory chapter will demonstrate how successive UK governments have gradually reduced the opportunities for nurturing calculation fluency in our primary classrooms. It also explains the reasons for identifying gender, prior attainment and confidence as the three main predictor variables in this calculation study. It begins by highlighting the increasing importance of nurturing fluency in our mathematics classrooms.

Our increasingly automated environment.

Today’s learners are growing up in an increasing automated environment. US-based researchers have called for schools to nurture mathematical ‘habits of mind’ to prepare learners for an unpredictable job market (Cuoco, Goldenberg & Mark, 1996, pp. 378-383). Since it has been predicted that up to 30% of current jobs are judged to be at high risk of automation by the early 2030s (Berriman, Hawksworth, Kelly & Foyster, 2017, p. 32), the ability to solve the types of non-routine problems which robots have not yet been able to address will become increasingly important for future generations across the globe. However, having the ability to solve a routine problem by following a taught procedure does not automatically translate into being able to solve non-routine problems too. For example, an international comparison study exploring problem-solving (Cai & Hwang, 2002) reported that Chinese learners outperformed their American counterparts in both their calculation and routine problem-solving tasks, but the results were reversed for non-routine problems where the Americans outperformed the Chinese. For example, the Chinese learners were able to perform long division (the formal algorithm for division calculations) but they struggled to round their answers correctly in real-life contexts.

The curriculum reforms.

Number and calculation skills are generally regarded as crucial aspects of many mathematics curricula. Focusing on the UK, successive governments have revised the statutory and advisory content of its national curriculum ever since its introduction under the Education Reform Act 1988. Indeed, the National Strategies (1999 – 2011) were established by the Department for Education and Employment (DfEE) to support schools adapt to further reforms across the curriculum. For mathematics, those reforms (DfEE, 1999, p. 70) instructed schools to encourage their learners to consider calculating mentally before attempting either a pencil-and-paper strategy or using a calculator. Several years later, the Department for Education and Science (DfES) replaced the DfEE. It issued guidance to schools promoting the use of 'efficient' pencil-and-paper strategies without specifying which strategies were deemed 'efficient' (DfES, 2007, p. 1). More recently, Education Minister Liz Truss MP (DfE, 2013b, para. 40) ushered in the most recent reforms which prioritised "efficient calculation methods like columnar addition and subtraction and short and long multiplication [the formal algorithm for multiplication]". Hence, 25 years after the introduction of the Education Reform Act, schools were no longer permitted the freedom to choose which calculation strategies they taught to their learners. Instead, they were compelled to teach formal algorithms for each of the four operations (Figure 1). Truss (2013b, para. 41-43) justified favouring formal algorithms by heavily criticising other calculation strategies, singling out the grid method and chunking (Figure 2) as "tortured techniques" which confused older family members. Following the 2013 reforms, calculators were no longer permitted in the SATs for Y6 learners (STA, 2016b, p. 2).

My own analysis revealed that the proportion of SAT marks awarded for number and calculation increased by over 13% following the reforms[1].

x	200	40	6
30	6000	1200	180
7	1400	280	42

6000+1200+180 = 7380
1400+280+42 = 1722
Answer 246x37 = 9102

```
       24
16 ) 384
   − 320     (20x16)
      64
    − 64     (4x16)
       0
```

Figure 2. Examples of the grid method and chunking. Adapted from *Good practice in primary mathematics: Evidence from successful schools* (2011, pp. 8-16) by the Office for Standards in Education, Children's Services and Skills (Ofsted).

Mathematics is a subject which is not taught in isolation. Cultural issues must also be considered. Learners are influenced by their families, their local community and increasingly by social media too. Hence the next sections will focus on the three predictor variables for this calculation fluency study, beginning with gender and prior attainment.

Gender and prior attainment.

Both gender and prior attainment were key variables in the theoretical model proposed by US-based Villalobos (2009). She argued that differences in gender and mathematical attainment observed in US school mathematics could be explained by the different curricula that learners experienced in their primary and secondary schooling. Although her model will be explored in more depth in the literature review, for this introductory section it is sufficient to note that her model predicted that the initial success that many US girls experienced with their number and calculation work at primary level hampered their transition to a more problem-solving dominated curriculum at secondary level.

The data indicated an existing gender difference in the UK for mathematical attainment which already widened as learners progressed through their schooling. Bearing in mind the changes arising in English schools due to the reforms (DfE, 2013a), Villalobos' model predicted that girls might respond positively to the increased focus on formal algorithms, yet their early success might hamper their later mathematical attainment at secondary level. Hence, it has become arguably even more important to consider the response of the learners to the recent increased focus on formal algorithms. The following paragraphs will explore the existing gender differences in more depth, beginning at primary level.

[1] Each question on the 2015 (pre-reform) and 2016 (post-reform) papers was coded and categorised into one of four strands: number and calculation; shape, space and measures; geometry; and, statistics.

For primary-aged learners, national assessments for Y6s indicated significant gender differences in performance at the higher levels; similar numbers of girls and boys consistently achieved the age-related Level Four (L4) expectation, but boys were outperforming girls at the higher Level Five (L5) (Figure 3).

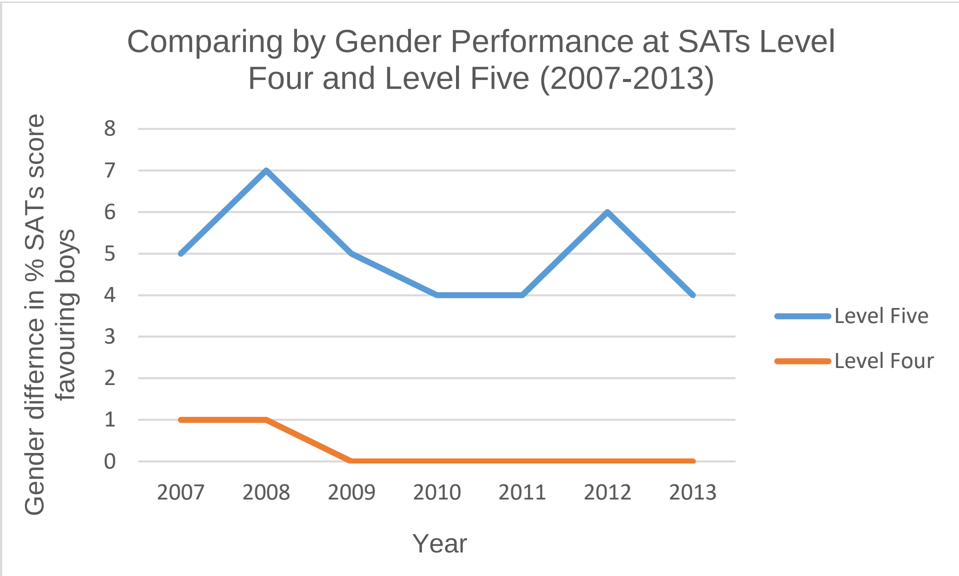


Figure 3. Comparing by gender performance at SATS Level Four and Level Five (2007-2013). Adapted from *National curriculum assessments at key stage 2 in England 2013* by DfE, 2013c. Retrieved from https://docs.google.com/spreadsheets/d/1vlmVznYbdn0vB47OHOFIm0BhZaDbQW_vINfe1xqZr5g/edit#gid=0.

Examining the assessment descriptors issued by the Department for Children, Schools and Families (DCSF, 2009), under the previous government's National Strategies initiative, revealed that the L4 descriptors for number and calculation focused on number recall and performing written calculations whereas those for L5 focused more heavily on problem-solving skills with numbers (Appendix B). In other words, L4 appeared to focus on procedural fluency whereas L5, with its focus on problem-solving which often required learners to work flexibly, was more closely aligned with the concept of calculation fluency discussed in the literature review. Since the data showed that boys tended to outperform girls at L5, it appeared that girls were less successful than boys at managing that change in expectations for problem-solving. Gender differences were also apparent in the General Certificate of Secondary Education (GCSE) mathematics results taken teenagers. Gill (2015, pp. 159-163) showed that boys consistently outperformed girls at the higher A* grade (Figure 4). Gender differences were not restricted to the UK; the literature also revealed gender differences in mathematical performance at international level too. The Programme for International Student Assessment (PISA) tests conducted by the Organisation for Economic Cooperation (OECD, 2014, p. 9) reported marked gender differences in mathematics performance among 15-year-olds favouring boys in many countries and economies.

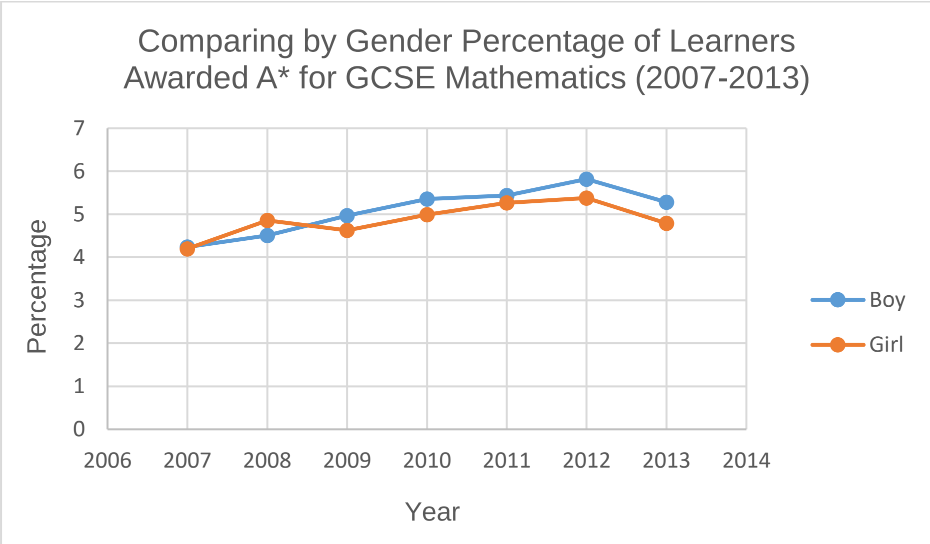


Figure 4. Comparing by gender percentage of learners awarded A* for GCSE mathematics (2007 – 2013). Adapted from "*GCSE uptake and results by gender 2004-2013*" by T. Gill, 2015, pp. 159-163. Copyright 2015 by Cambridge Assessment. Adapted with permission.

Encouraging learners of both genders to study mathematics post-16 addresses social justice concerns as well as ensuring a steady supply of STEM-trained students and employees; yet the OECD (2012, p.1) reported that most countries demonstrated a statistically significant gender difference favouring boys in the proportion of 15-year-olds planning a career in engineering or computing. Focusing on the UK, their report revealed that more than six times the number of 15-year-old boys than girls were considering a STEM-related career (OECD, 2012, p. 2).

The gender differences in mathematics favouring boys seemed to widen even further at university level (McWhinnie & Fox, 2013). Apart from a slight increase at undergraduate level, which was still below the proportion of male undergraduates studying mathematics, Figure 5 clearly illustrated the low proportion of female mathematicians working in English universities.

Hence gender differences, noticeable in overall mathematical performance among higher attaining primary-aged learners, appeared to widen as learners progressed through their schooling and embarked on their future careers or further studies. If Villalobos' (2009) model was correct, then the recent reforms might extend that gender gap even further. Hence both gender and prior attainment were two of the three predictor variables in this calculation fluency study.

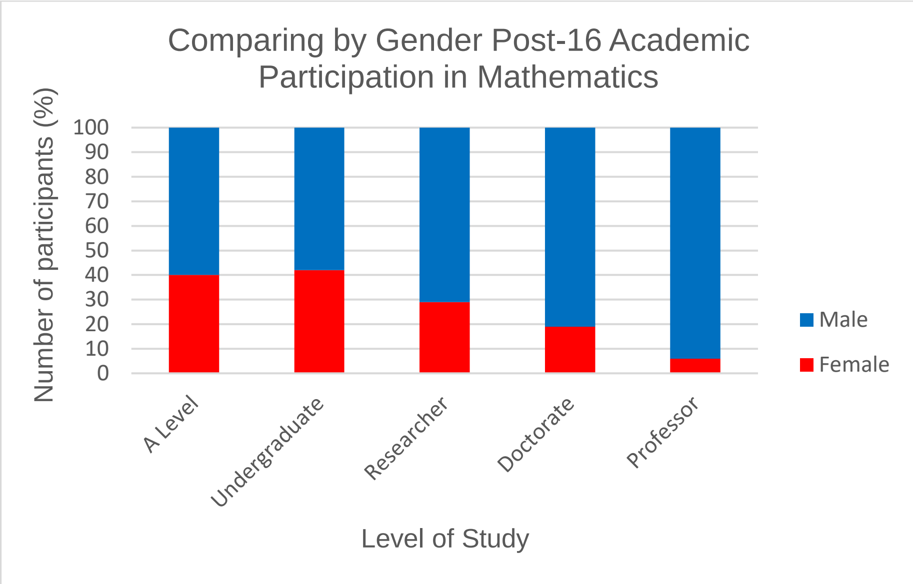


Figure 5. Comparing by gender post-16 academic participation in mathematics. Adapted from *Advancing women in mathematics: Good practice in university departments* by S. McWhinnie and C. Fox, 2013. Copyright London Mathematical Society. Adapted with permission.

Confidence.

Mathematical confidence was the third predictor variable for this study. Of particular relevance for this study were the findings of Nunes, Bryant, Sylva and Barros (2009, p. 5) who reported significant gender differences in mathematical confidence favouring boys among English 8- and 9-year-olds ($N \approx$ 4,000). At that age, those learners were experiencing a curriculum dominated by number and calculation. Hence a possible interpretation of their results was that the girls were less confident about their number and calculation skills than the boys. Moreover, other studies (including Brown, Brown, & Bibby, 2008; Heilbronner, 2013; Pampaka, Kleanthous, Hutcheson, & Wake, 2011) indicated that confident students were more likely to continue studying mathematics post-16 than their less confident counterparts. More specifically, US-based Heilbronner (2012) explored the factors surrounding gender participation in the various STEM-related career paths. Although her results indicated that roughly equal numbers from both genders entered STEM-related careers in the US, she also reported that less confident female students were less likely to choose engineering and mathematics than their male counterparts. In other words, there was a gender imbalance across the STEM-related subjects. Also, Brown et al. (2008, p. 12) reported that almost twice as many UK-based 16-year-old girls than boys indicated that they would not consider studying mathematics post- 16 because it was perceived as ‘too difficult.’ Focusing on UK students who had elected to study mathematics post-16, Pampaka et al. (2011) reported that their mathematical confidence was related to both their attainment and their gender. However, even though this paragraph indicates the importance of confidence, the available literature addressing calculation and confidence was scarce.

Thesis Aims and Objectives

This thesis addressed calculation fluency during a period of curriculum change in English primary schools. Although calculation was a theme which featured heavily in the literature, the review in the next chapter will show that the predominance of choice/no-choice research designs (including Carr & Davis, 2001; Torbeyns & Verschaffel, 2013) provided limited evidence regarding calculation flexibility. Moreover, the literature will also reveal that none of the available studies explored the reasons behind learners' calculation choices with the learners themselves. As far as the four operations were concerned, the literature for the younger learners tended to focus on addition or subtraction, or both (including Carr & Alexeev, 2011; Fennema, Carpenter, Jacobs, & Levi, 1998), whereas the studies conducted with older learners often focused on division (including Anghileri, Beishuizen, van Putten & Snijders, 1999; Hickendorff, van Putten, Verhelst & Heiser, 2010). A longitudinal study exploring calculation strategies across all of the four operations, conducted by Borthwick and Harcourt-Heath (2007, 2010, 2012, 2015 and 2016), was arguably hampered by its reliance on secondary data consisting of a single calculation for each of the four operations. Taking into account the above gaps in the evidence, alongside the recent reforms (DfE, 2013a) prioritising formal algorithms and the growing need for adaptability in the workplace, the focus of this thesis on calculation fluency appeared timely. In order to address the gaps in the literature, this study collected both quantitative and qualitative data. Hence my decision to adopt a mixed methods research (MMR) design, which is described in detail in the methodology chapter, involving a large-scale survey followed by a smaller number of follow-up interviews.

In the opening sentences of this introductory chapter, mathematics was portrayed as a creative, flexible discipline with the potential to inspire learners in both their studies and everyday lives. Moreover, ensuring that future generations develop the adaptable skills which will enable them to thrive in our increasingly automated environment is becoming crucially important. Nurturing fluency is perhaps more important than ever before. However, recent curriculum reforms in the UK (DfE, 2013a) appeared to prioritise the teaching of formal algorithms, arguably reducing the opportunities for young learners to work flexibly with numbers and calculations, and possibly hindering the development of their calculation fluency of both genders but most especially girls. Hence this MMR study, which focused on the first cohort of Y6 learners studying under the reforms, addressed the following overarching question:

> ‘To what extent does calculation fluency among Y6 learners vary by gender, confidence level and prior attainment?’

The four key aspects of calculation fluency identified during the literature review and the three variables introduced in this chapter influenced the design of following four subsidiary questions:

RQ1: To what extent do gender, prior attainment or confidence predict use of the formal algorithm?

RQ2: To what extent do gender, prior attainment and confidence predict calculation accuracy?

RQ3: Which are the most accurate calculation strategies for Y6 learners completing age-related, context-free multi-digit written calculations?

RQ4: To what extent do calculation efficiency and understanding vary by gender, prior attainment and confidence?

Outline of Thesis Chapters

This introductory chapter is followed by a detailed literature review in Chapter Two which will examine in more depth the literature surrounding calculation fluency, gender, prior attainment and confidence. The review will examine the research designs and findings of prior studies to identify gaps in the literature and inform the design of this study. Chapter Three will present the philosophical assumptions behind this thesis and the decision-making process which resulted in the choice of an MMR design. It will explain how the literature review findings influenced the design of the data collection instruments for this study and its sampling approach too. The steps taken to ensure that this study was conducted in an ethical manner will also be outlined in the methodology chapter. Chapters Four and Five will present the findings deriving from analysis of the quantitative and qualitative data, including the integration of the findings from both phases of the study. This thesis will conclude with a detailed discussion of those findings in Chapter Six in relation to the four subsidiary RQs and overarching RQ, followed by consideration of the implications for researchers as well as schools and policymakers.

Chapter Two: Literature Review

Introduction

This review considers the existing literature relating to the theme of calculation fluency. It begins by detailing the approach taken towards the literature review, including the identification of the search terms as well as the inclusion and exclusion criteria. This is followed by a discussion of the literature relating to the key aspects of calculation fluency, beginning with the literature addressing the relationship between calculation and number sense, including the relevant evidence surrounding learners experiencing mathematical learning difficulties. The review continues by exploring the literature relating to the three factors introduced in the previous chapter, namely mathematical confidence, gender and prior attainment. This is followed by a discussion concerning the literature relating to the relative importance of acquiring calculation fluency in the digital era. The review concludes by addressing the anticipated contribution of this thesis.

Approach Taken Towards Literature Review

Search terms.

The process of identifying the initial search terms began by reviewing the overarching RQ for this thesis: "To what extent does calculation fluency among Y6 learners vary by gender, confidence level and prior attainment?" Hence *calculation*, *gender difference* and *confidence* were identified as key words for the literature search and synonyms for each of those words were identified using the thesauri of the British Education Index (BEI) and Educational Resources Information Center (ERIC). This process resulted in the full list of initial search terms presented in Table 1.

Table 1

The Initial List of Search Terms

Initial Search Terms		
Calculation	Gender Difference	Confidence
Algorithm	Sex difference	Self-esteem
Arithmetic		
Computation		

Selection of the databases and other sources.

The next stage of the literature review process involved using the initial list of search terms to check the databases of the BEI, Child Development & Adolescent Studies, ERIC, Education Abstracts, JSTOR, PsysARTICLES, PsycINFO and the Teacher Research Center as well as Google and Google

Scholar. The review also searched the catalogues of the peer-reviewed mathematics journals Research in Mathematics Education and Educational Studies in Mathematics, as well as the Proceedings of the British Society for Research into Learning Mathematics. Other searches included UK government publications, the UK-based subject association websites for the Association of Teachers of Mathematics (ATM) and the Mathematical Association (MA), the websites of the US-based National Council of Teachers of Mathematics (NCTM) and UK-based National Centre for Excellence in the Teaching of Mathematics (NCETM) and the British Library's E-theses Online Service (EThOS).

Screening process.

The initial search revealed a wealth of available literature and a screening procedure identified the most relevant studies. This screening process excluded:

- studies published in languages other than English.
- studies conducted prior to 1970.

The initial search raised the importance of both *number sense* and *mathematical learning difficulties* as two additional themes for this review. Hence an additional search was conducted, following the same approach as the initial search, exploring the literature addressing those two additional themes as the search terms. The remainder of this chapter presents the findings from the review.

Calculation

Mathematical learning difficulties and number sense.

This review begins by considering the literature addressing the difficulties faced by some learners acquiring number skills and its impact on my research design. Although the evidence indicated an overall improvement in Y6 SATs performance in the years following the introduction of the National Strategies, "differences in arithmetic are very marked" (Dowker, 2004, p. 5). This was not a new situation: Cockcroft (1982, para. 342) reported a wide range of calculation skills among primary-aged learners and suggested that a typical Y6 class might include learners with a 7-year difference in their calculation skills. More recently, Geary, Hoard, Nugent and Bailey (2012) reported that around 7% of learners had a *mathematical learning disability* (MLD). While it should be noted that MLD refers to the wider mathematics curriculum, the literature frequently referred to a specific learning difficulty known as *dyscalculia*. When reviewing the available literature for dyscalculia, it should also be noted that its study was very much in its infancy compared to the much wider body of evidence relating to dyslexia (Berch & Mazzoco, 2007). Attempting to define dyscalculia, Gross (2007, p. 152) argued that educators should adopt its literal meaning as "the inability to calculate." Although her interpretation of dyscalculia highlighted the relevance of considering dyscalculia in my research design, her definition implied that all learners who struggled with their calculation skills were to a certain extent dyscalculic. Such a definition seemed exceptionally broad since many, if not all, learners, face difficulties at some points when learning to calculate. In contrast, Canadian researcher Ansari (2014) adopted the terminology *developmental dyscalculia* to describe learners facing persistent difficulties in their number work, further noting that that the condition affected approximately 5% of learners.

These findings relating to dyscalculia had implications for my own sampling decisions. More specifically, it did not appear appropriate to expect dyscalculic learners to participate in a calculation study unless it focused on their needs. Since the above literature indicated that the proportion of dyscalculic learners in a mainstream classroom was possibly as low as one or two learners per average-sized UK class (Ansari, 2014), then excluding those learners from my study should not significantly affect my data collection or the generalisation of my findings.

Number sense.

The literature relating to MLDs frequently referred to the development of *number sense* and calculation skills. For example, Dowker (1992) stated that:

> To the person without number sense, arithmetic is a bewildering territory in which any deviation from the known path may rapidly lead to being totally lost. The person with number sense… has, metaphorically, an effective "cognitive map" of that same territory, which means that such deviations can be tolerated, since the person can expect to be able to correct them if they cause problems and is unlikely to become lost in any serious sense. (p. 52)

Nevertheless, any attempts to clarify the meaning of number sense and its relationship with the development of calculation skills were hampered by the realisation that “no two researchers in the field accept the same definition” (Back, 2014, para. 3). Indeed, Case (1998) cautioned that “Number sense is difficult to define but easy to recognize” (p. 1). Those researchers who offered a definition number sense in their papers tended to refer to the following:

> Number sense refers to a person's general understanding of number and operations along with the ability and inclination to use this understanding in flexible ways to make mathematical judgements and to develop useful strategies for handling numbers and operations. It reflects an inclination and an ability to use numbers and quantitative methods as a means of communicating, processing and interpreting information. It results in an expectation that numbers are useful and that mathematics has a certain regularity (McIntosh, Reys & Reys, 1992, p. 3).

This very broad definition encompassed both number and calculation, requiring learners to demonstrate their understanding as well as make decisions about their calculation strategies. Moreover, flexibility appeared to be an essential aspect of number sense, but not necessarily knowledge of formal algorithms. By stressing the need for "useful strategies," McIntosh et al. possibly implied the need for efficiency too.

Although such a broad definition of number sense was challenged in the literature because it "includes…most, if not all, other skills and dispositions…related to number and arithmetic" (Verschaffel, Greer & De Corte, 2007, p. 558), the importance of flexibility appeared to be a recurring theme. For example, in their paper comparing the benefits of developing phonic awareness among learners facing reading difficulties with developing number sense for learners struggling with their early mathematics, Gersten and Chard (1999, p. 19) called for a greater focus on "a child's fluidity and flexibility with number."

Meanwhile Berch (2005) suggested that number sense operated at two levels; a lower-level for counting and performing simple arithmetical operations and a higher-level demanding "a high degree of fluency and flexibility with operations and procedures" (p. 334). If learners did indeed move from one level of number sense to another, then it appeared that number sense is not fixed. An alternative interpretation was suggested by Anghileri (2000) who proposed that it "develops continually as the range of known facts and the relationships among them are extended" (p. 6). She drew attention to the close relationship between number sense and calculation, "Although children may learn some standard procedures, using number sense involves departure from these methods where the numbers warrant a different approach" (p. 127). This was a key point for my research design; if number sense required learners to make decisions about their calculation choices, then my study design required sufficient calculations to enable them to demonstrate those skills.

Reflecting on the above discussion, the definition of number sense adopted for this study required learners to demonstrate their understanding about numbers, and the relationships between them, as well as show a willingness to apply their knowledge in a flexible and efficient manner where appropriate. The following sections in this review involve exploring the evidence addressing the relationship between number sense and calculation fluency, beginning with the literature addressing the meaning of calculation fluency.

Calculation fluency.

The terms *calculation* and *computation* appeared to be synonymous in the literature. However, Google's Ngram Viewer (Books.google.com, 2017) revealed that the term calculation occured more frequently in its corpus than computation. Hence calculation was the adopted terminology for this thesis.

The literature addressing fluency was less straightforward because it revealed a plethora of relevant terms which included *calculation fluency*, *procedural fluency*, *mathematical fluency* and *algorithmic fluency*. It could be argued that procedural fluency was not necessarily specific to the number and calculation strand of a curriculum; it could equally refer to other aspects of mathematics, such as following instructions to bisect an angle in a geometry session. A similar argument applied to both algorithmic fluency and mathematical fluency. In contrast, calculation fluency specifically focused on the curriculum area of calculation, making it the most relevant terminology for this study.

Although the literature revealed a large volume of studies relating to calculation, few of those studies specifically addressed calculation fluency. This omission indicated that a possible contribution of this thesis was its potential for extending our knowledge regarding calculation fluency. However, it also reduced the opportunities for reflecting on existing definitions of calculation fluency. In one of the scarce academic studies offering a definition of calculation fluency, Baroody, Torbeyns and Verschaffel (2009) suggested that it required "the efficient, appropriate, and flexible application of arithmetic skills" (p. 3). Reflecting on the previous section addressing number sense, Baroody et al.'s definition of calculation fluency appeared to fall within the higher-level of number sense (Berch, 2005) or towards the far end of a number sense continuum (Anghileri, 2000). Hence, assuming that number sense was a prerequisite for calculation fluency, combining Baroody et al's (2009) definition highlighting the importance of flexibility, efficiency and decision-making with the number sense definition adopted for this study in the previous section, indicated that calculation fluency required flexibility (Anghileri, 2000; Baroody et al., 2009; Berch, 2005; Gersten & Chard, 1999; McIntosh et al, 1992), efficiency (Baroody et al, 2009) and understanding (McIntosh et al, 1992). Moreover, I suggest that accuracy was implied by Berch's (2005) reference to "performing simple operations" (p. 334) and Baroody et al.'s (2009, p. 3) "application of arithmetic skills." Hence this study defined calculation fluency as requiring learners to demonstrate their understanding, flexibility, efficiency and accuracy when performing calculations.

The next four sections of this chapter explore the literature relating to each of those four key aspects of calculation fluency in more depth, beginning with flexibility. It should be noted that the review also revealed the inter-related nature of those four key aspects; their close association is illustrated by Figure 6.

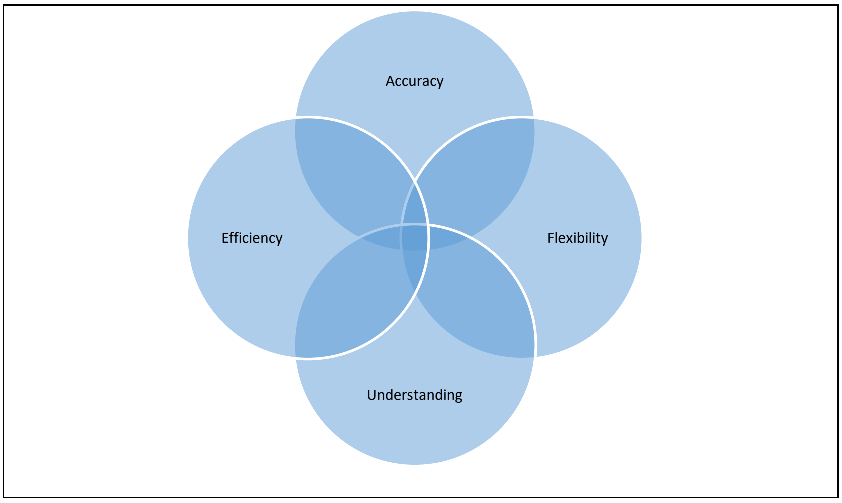


Figure 6. Diagram illustrating the inter-connected nature of the four aspects of calculation fluency (Source: Personal collection).

Flexibility.

The perceived importance of calculation flexibility was a key motivation for this thesis which followed the implementation of UK government reforms prioritising formal algorithms over and above other strategies (DfE, 2013a). The literature highlighted the benefits of being able to calculate flexibly (Ma, 1999):

> Being able to calculate in multiple ways means that one has transcended the formality of the algorithm and reached the essence of the numerical operations - the underlying mathematical ideas and principles. The reason that one problem can be solved in multiple ways is that mathematics does not consist of isolated rules but connected ideas. Being able to and tending to solve a problem in more than one way, therefore, reveals the ability and the predilection to make connections between and among mathematical areas and topics. (p. 112)

Although Ma championed working flexibly, she also noted that learners needed to develop an inclination to use such skills. A similar theme arose in the earlier discussion defining number sense (McIntosh et al., 1992, p. 3). Likewise, Star, Rittle-Johnson and Durkin (2016) called for learners to know multiple strategies and demonstrate the critical judgement to choose between them for particular problems. Since my study was based in a jurisdiction prioritising formal algorithms (DfE, 2013a), my interpretation of flexibility was driven by local circumstances; learners were deemed to be working flexibly if they deviated from using a formal algorithm for a multi-digit calculation. Nevertheless, since the above literature highlighted the importance of working flexibly when appropriate to do so, my research design also needed to address the development of a range of multi-digit calculations deemed suitable for alternative strategies to the formal algorithm by educational professionals.

At first glance, the literature appeared to reveal a shifting picture regarding flexibility in the UK classroom. The predominance of formal algorithms 20 years ago (Anghileri, 2001) decreased in the years following the introduction of the National Strategies (Anghieri, 2006; Borthwick & Harcourt-Heath, 2007). More recently, the situation reversed again as formal algorithms returned to prominence (Borthwick & Harcourt-Heath, 2016). Nevertheless, I suggest that these findings did not necessarily imply that the flexibility of learners was changing over time because the research designs of those studies focused on the number of times that their overall sample chose the formal algorithm rather than exploring the ways in which an individual learner might vary their calculation choices over a range of calculations. It appeared that there was a gap in the available literature regarding the willingness of school-age learners to adapt their strategy choices according to the numbers in an individual calculation; existing calculation studies often adopted a choice/no-choice data collection approach (including Carr & Davis, 2001 and Torbeyns & Verschaffel, 2013, which I shall explore in more depth later in this chapter) but the key point for this aspect of the review was that such an approach necessarily limited the range of calculation choices available for their participants.

The literature did address calculation flexibility with adults and it revealed that having a choice between working mentally, using a written strategy or a calculator often resulted in better overall calculation performance among adults (Siegler & Lemaire, 1997, p. 72). The researchers also reported a preference among their adult participants for mental rather than pencil-and-paper calculations (p. 88). Moreover, the literature showed that adults often adopted idiosyncratic methods rather than formal algorithms (Cockcroft, 1982, pp. 19-20). However, the available literature did not address the reasons behind those choices.

Reflecting on the literature regarding calculation flexibility, my research design could collect data regarding a potentially much wider range of calculation strategies than appeared in the existing literature. Second, the existing calculation studies discussed so far in this chapter tended to focus on collecting quantitative data rather than examining the reasons behind the calculation choices of the participants through interviews, including their views on working flexibly. By designing a study which included also a qualitative aspect to data collection, my research design could help to address that gap in the literature. Nevertheless, all research designs needred to make compromises. By interpreting flexibility as deviating from the formal algorithm, I accepted that, say, a learner exclusively relying on the grid method for multiplication might have appeared to be working flexibly even though they might have actually been inflexible about its use.

Efficiency.

The literature suggested that learners demonstrated efficient working when:

> The learner does not get bogged down in many steps or lose track of the logic of the strategy. An efficient strategy is one that the learner can carry out easily, keeping track of subproblems and making use of intermediate results to solve the problem (Russell, 2000, p. 5)

Russell made a crucial point when she suggested that efficient strategies were ones that that learners could carry out easily. Moreover, Anghileri (2000) suggested that "if the purpose is to find a solution to a particular problem using an efficient and reliable method, then individual children should be encouraged to use any method they are confident satisfies this requirement" (pp. 97-98). However, since the perception of an easy task could vary from learner to learner, the perceived easiness of a strategy appeared to be more amenable to qualitative rather than qualitative data collection approaches, and this was a key point for my research design.

Russell's statement also drew attention to the relationship between efficiency and the number of steps in an algorithm. She drew attention to the calculation 1002 – 998 where counting-up, rather than executing the formal algorithm, might have been more efficient for many learners. Hence formal algorithms were not necessarily the most efficient strategy for all multi-digit calculations because mental approaches might have been faster and less error-prone for some calculations. Indeed, Russell argued that if a 10-year-old learner attempted the formal algorithm for such a calculation then "it is time to worry about what the learner knows about whole number operations" (2000, p. 2). This example again highlighted the importance of understanding the reasons behind the calculation choices of learners; an individual's strategy might have appeared inefficient, but learners developed their calculation skills at different rates and an efficient strategy for one learner might have seemed inefficient for another one. Evidently, there appeared to be a close relationship between flexibility and efficiency; by working flexibly learners might be able to calculate more efficiently than relying on a single strategy for all calculations. Nevertheless, a learner might have been working flexibly but not necessarily efficiently; knowing when to choose certain strategies was important too. "Children should be able to select and perform procedures based on an understanding of why and when these procedures are appropriate" (Gilmore & Bryant, 2008, p. 4). By adopting the definition of an efficient strategy for this thesis as "one that the learner can carry out easily" (Russell, 2000, p. 5), learners were placed at the centre of the process of evaluating the efficiency of their calculation strategy.

Although it was not necessarily correct that formal algorithms were the most efficient strategies for all multi-digit calculations, the reformed curriculum (DfE, 2013a) arguably reinforced the widely-held view that formal algorithms were more socially-acceptable than other calculation strategies (Cockcroft, 1982, p. 8; Torbeyns & Verschaffel, 2013, p. 114). For example, Russell (2000, p. 1) recalled a Year 2 (Y2) boy incorrectly calculating 57 x 4 using the formal algorithm. Her subsequent discussion revealed that the boy was capable of correctly calculating the answer by partitioning into 50 x 4 and 7 x 4, and she concluded that “for this learner, 57 x 4 should have been an easy mental problem.” However, she discovered that he was “very proud” of knowing the formal algorithm. Moreover, Fielker (2007) highlighted the example of a high-attaining 13-year-old boy who performed the calculation 20 x 36 using the formal algorithm even though he admitted that he could have easily calculated the answer mentally. By appearing to prioritise a strategy over and above the numbers in a calculation, he arguably completed a calculation less efficiently than might have been the case.

Reflecting on the literature surrounding calculation efficiency, it appeared that individuals might have held different views regarding the relative efficiency of a calculation strategy. Therefore, my data collection needed to consider the views of the learners regarding the efficiency of their calculation choices.

Accuracy.

This section examines the literature regarding the most accurate strategies for different operations as well as considering the literature addressing estimation, checking the correct application of procedures and reflecting on the reasonableness of answers. For the purposes of this study, an accurate calculation was defined as a calculation yielding the correct answer.

However, I agreed with Russell (2000) who argued that calculation accuracy “depends on several aspects of the problem-solving process, among them, careful recording, the knowledge of basic number combinations and other important number relationships, and concern for double-checking results” (p. 5). Hence the next part of the review will address the literature relating to checking procedures as well as findings regarding the most accurate strategies.

The most accurate strategies.

Although the DfE’s reforms (2013a) prioritised formal algorithms, the available literature appeared to cast doubt on any assumptions that they were the most accurate strategies for all multi-digit calculations. Based on a study by Borthwick and Harcourt-Heath (2007) it appeared that more Y5 learners were successful subtracting with a number line and dividing by chunking down than executing the relevant formal algorithm (N = 973). However, their study’s research design was restricted by its use of secondary data taken from completed 2005 SAT papers which resulted in the authors relying on a single calculation for each of the four operations (546 + 423, 317 - 180, 56 x 24 and 222 ÷ 3). Moreover, Borthwick and Harcourt-Heath (2007) acknowledged that the addition calculation was below the level of challenge posed by the calculations for the other three operations. There appeared to be a scarcity of literature addressing the accuracy rates of strategies chosen by individual learners completing a variety of age-related, multi-digit calculations; the predominance of the choice/no-choice research design for calculation studies limited further our ability to compare the accuracy rates of a wide range of strategies. Hence, if my research design enabled the comparison of accuracy rates across a range of age-related calculations for each of the four operations, then my findings might help to address that gap in the literature.

Following up their initial study, Borthwick and Harcourt-Heath (2015) reported improvements in calculation accuracy across all the four operations between 2006 and 2014 (Table 2) and an increase in the use of number lines for both subtraction and division, and it should be noted that these findings related to learners unaffected by the recent reforms (DfE, 2013a). Borthwick and Harcourt-Heath (2015) also raised concerns that the proportion of learners not attempting the multiplication or division calculations remained static at around a fifth of their sample. Hence my research analysis could consider the proportion of learners who did attempt those calculations following the implementation of the reformed curriculum (DfE, 2013a).

Table 2

Changes in Calculation Accuracy between 2006 and 2014 (*N* = 1,021)

Calculation	% Change
546 + 423	+2
317 – 180	+33
56 x 24	+29
222 ÷ 3	+26

Note. Adapted from *Calculating: How have Year 5 children's strategies changed over time*? (pp. 2-5) by A. Borthwick and M. Harcourt-Heath, 2015. Copyright 2015 by Alison Borthwick and Micky Harcourt-Heath. Adapted with permission.

Clearly, many learners were either making errors or not attempting certain calculations; over half of the learners in the final year of Borthwick and Harcourt-Heath's longitudinal study were unable to complete the division calculation (Table 3) (Borthwick & Harcourt-Heath, 2015).

Table 3

Calculation Accuracy across the Four Operations (*N* = 1,021)

Calculation	% correct answers	% incorrect answers
546 + 423	92	8
317 - 180	75	25
56 x 24	51	49
222 ÷ 3	47	53

Note. Adapted from *Calculating: How have Year 5 children's strategies changed over time*? (pp 2-5) by A. Borthwick and M. Harcourt-Heath, 2015. Copyright by Alison Borthwick and Micky Harcourt-Heath. Adapted with permission.

This finding brought the discussion to another issue facing calculation studies - consistency in accuracy over time. For the age-related subtraction calculation 82 – 3.8, slightly older 13-to-14-year-old learners using a number line "were more successful, whereas those selecting the decomposition standard algorithm appear not to have had the understanding to enable them to utilise this effectively" (Borthwick, Harcourt-Heath & Keating, 2014, p. 28). However, although it would have been interesting to track the calculation choices of learners over several years to explore those findings in more depth, it was beyond the scope of my own research design.

The literature also addressed gender and accuracy rates, which was highly relevant for my research design. Among 6- to 8-year-olds (N = 82) performing addition and subtraction calculations, Fennema, Carpenter, Jacobs and Levi (1998) reported no gender differences in calculation accuracy. However, they did indicate gender differences in strategy choices and those findings are explored in more depth later in this chapter. Meanwhile the literature also revealed a longitudinal study among primary-aged learners (N = 311) which reported gender differences in attitudes towards calculation accuracy whereby girls tended to prioritise accuracy over efficiency, preferring counting with their fingers over working mentally (Bailey, Littlefield & Geary., 2012). My research design could build on those findings by exploring the approaches taken towards ensuring calculation accuracy during the qualitative data collection.

The literature showed that different strategies could result in different errors. For example, typical learner errors using the grid method were mistakes calculating partial products and totalling those partial products to reach the final answer: "children recorded an answer of 100 for the partial product of 50 x 20" (Borthwick & Harcourt-Heath, 2007, p. 3). In contrast, a common long multiplication error was the incorrect use of zero as a place-holder (Ofsted, 2011, p. 28). US-based Keiser (2012, p. 412) argued that teachers should accept that different strategies would lead to different errors and plan their lessons accordingly. Moreover, Ofsted (2011) recognised the benefits of using the grid method in secondary schooling, noting that it tended to reduce errors when learners applied that strategy to the algebraic expansion of brackets such as $(2x + 3)\ (x – 6)$:

> It is particularly valuable in emphasising the four products, thereby tackling the common error where only the first and last terms in each bracket are multiplied... It also provides insight into the reverse process, factorisation, which learners generally find more difficult. (p. 8)

Nevertheless, although the inspectors acknowledged the future benefits of using the grid method for multiplication, they also criticised using the chunking down division strategy with larger numbers because "such an approach leads to errors" (Ofsted, 2011, p. 16). The inspectors highlighted the difficulties faced by some learners attempting chunking when taking away small chunks several times because they did not appear to spot the opportunity to remove larger chunks. The inspectors concluded that learners should be encouraged to progress to the formal algorithm for division rather than attempt chunking down with larger numbers. The literature revealed disagreement with such comments:

> In fact, the strength of the chunking algorithm lies in its great potential for differentiation: it allows for a range of levels of sophistication in children's confidence and understanding, in that the less confident can remove small chunks; the more confident can take away larger chunks; and the most confident can subtract the maximum-sized chunks. (Thompson, 2008, p. 7)

While Thompson suggested differentiating a single strategy, Star (2005) noted that knowing more than one strategy enabled learners to identify a less error-prone method for a calculation rather than relying on a single strategy for all situations when working in situations beyond simple arithmetic such as multi-digit calculations.

Reflecting on the literature addressing the most accurate calculation strategies revealed the scarcity of studies exploring strategy choices across a range of calculations and operations. My research design could help to address this gap in the literature by including a range of age-related multi-digit calculations across the four operations. Moreover, exploring approaches towards ensuring calculation accuracy with the learners themselves might shed new light on their decision-making.

Checking procedures.

We all make mistakes, but checking our answers enables our mistakes to be identified, rectified and hence improve our calculation accuracy. However, the review revealed that few studies addressing checking procedures had been published in recent years. Dowker (1992) addressed errors in mathematics, drawing attention to the comments made by Hadamard (1945), "Good mathematicians, when they make them, which is not infrequent, soon perceive and correct them" (p. 49). Nevertheless, learners must first spot their mistakes if they are going to be able to correct them. Encouraging learners to work systematically and review their answers was promoted by McIntosh et al. (1992) who regarded such skills as evidence of number sense. Checking answers was also crucial for Russell (2000) who supported the 'double-checking' of results (p. 5).

Although reporting the total number of correct calculations by a learner indicated their overall accuracy, the statements by Dowker (1992) and Russell (2000) indicated that exploring the ways in which learners addressed checking their answers might be another aspect for consideration in my research design. However, written work in calculation workbooks might have provided insufficient evidence of exploring checking procedures. For example, some learners might have checked their answers by visually reviewing their working out whereas others might have tried adopting a different mental approach to check their initial written strategy. In both cases, there might have been no written evidence of their checking procedures. Also, Cockcroft (1982) recommended encouraging learners to apply other checking procedures which linked to their prior learning, such as knowing “that the sum of two odd numbers must be even or that any number in the '5 times table' ends in either 0 or 5” (para. 237) and those approaches might not have been obvious in any workbook recording.

Reflecting on the evidence addressing checking calculations revealed the scarcity of recent literature and the difficulties in collecting evidence from written work relating to checking procedures. My study could address these issues by exploring checking procedures with learners themselves as part of my interviews with them.

Reflecting on the reasonableness of an answer.

A limited number of studies in the literature raised the importance of checking the reasonableness of an answer. Consider the following: “Accuracy extends beyond just getting the correct answer. It involves considering the meaning of an operation, recording work carefully, and asking oneself whether the solution is reasonable” (Gojak, 2012, para. 3). By choosing the wording “asking oneself,” Gojak appeared to differentiate between checking that a procedure has been correctly applied and reflecting on the overall answer. The importance of reflecting on the reasonableness of an answer during place value work was also highlighted by Keiser (2010): “If I asked them to multiply or divide by a power of ten, they used the standard algorithm rather than simply moving the decimal point. They lacked the ability to determine if an answer was reasonable.” (p. 69). There appeared to be a subtle difference between checking a procedure and judging its reasonableness. For example, a learner might have incorrectly chosen to subtract when they should have multiplied; they might have applied their subtraction procedure correctly yet still reached the wrong answer. Such errors could often be spotted if learners reflected on the reasonableness of their answers rather than merely making procedural checks on their working out.

Reflecting on the literature relating to checking the reasonableness of an answer, although some researchers stressed its importance, there was an apparent lack of empirical studies in this area. Hence my study could help to address this gap in the literature by exploring during interviews whether and how learners checked the reasonableness of their answers.

Estimating.

There appeared to be a very close association between reflecting on the reasonableness of an answer and estimating since both approaches could be deployed to check an answer. However, estimating could also take place before performing the actual calculation whereas considering the reasonableness could only take place afterwards and could be used to reflect on the estimate.

The limited literature highlighted the importance of estimating skills for everyday life as well as classroom mathematics: “Industry and commerce rely extensively on the ability to estimate” (Cockcroft, 1982, para. 78). Yet, just two years later, Threadgill-Sowder (1984) reported that few adults estimated their answers and challenged the perceived assumption of the importance of estimation when their data revealed that it was rarely employed by adults. However, if estimation relied solely on mental calculation, then there might have been little evidence of it having taken place in a learner’s workbook and that would have had implications for my research design. The literature revealed a close association between estimation and calculation; the importance of nurturing estimating skills for improved calculation work was noted by Dowker (1998 and 2014).

Nevertheless, adopting any checking procedures such as estimating demanded additional time. It could be argued that the formal nature of SATs might increase the willingness of the Y6 learners in my study to estimate and check their work due to the perceived importance of the test situation, but the counter-argument might run that the time issue might make the learners less likely to devote time to estimating their answers. One possible approach suggested in the literature involved presenting learners with fewer calculations to ensure that they had had enough time to reflect on their approaches (Gojak, 2013, para. 6). Hence, I would need to balance the number of calculations in my study with allowing sufficient time for learners to complete any checking procedures.

Reflecting on the literature addressing estimating revealed similar issues as reported for checking the reasonableness of answers; researchers stressed the importance of estimating yet there was limited research addressing estimating skills among learners in the available studies. My research design could address this gap in the literature by drafting, piloting and refining a workbook which ensured that most learners had enough time to complete their calculations and those learners who did wish to check their calculations were usually able to do so.

To summarise this overall section addressing calculation accuracy, the literature review highlighted several gaps in research knowledge regarding the most accurate strategies and the checking procedures of learners. These gaps indicated that the research design for my own study would need to have both quantitative and qualitative components. The quantitative component would need to consider the total number of correct answers across the set of calculations covering the different operations to compare the accuracy rates of the different strategies chosen by the learners. The qualitative component would need to uncover whether learners estimated their answers, checked their working out and reflected on the reasonableness of their answers. The piloting of my data collection instruments would need to establish the length of time required for Y6 learners to complete the calculations and check their answers too.

Understanding.

According to the definition of calculation fluency adopted for this study, learners should understand their calculation strategies. However, the relative importance of understanding in the mathematics education literature has provoked heated debates which were retrospectively dubbed the "math wars" (Klein, 2007). Those debates, which arose in the US in the late 1980s, addressed the relative importance of conceptual and procedural knowledge. Hiebert and Lefevre (1986) defined conceptual knowledge as "knowledge that is rich in relationships" (p. 3) whereas they interpreted procedural knowledge as "rules or procedures for solving mathematical problems" (p. 7).

Of relevance to my focus on calculation fluency, the debates often focused their attention on the teaching of calculation to primary-aged learners; the traditionalists supported a curriculum focusing on procedural knowledge, arguing that schools should explicitly teach formal algorithms, whereas the reformers welcomed a reduced initial focus on learning formal methods in favour of nurturing conceptual knowledge. However, the perceived primacy of formal algorithms was challenged by US-based researchers McIntosh et al. (1992) who reported that adults made relatively little use of traditional approaches: "It is somewhat ironic that many people still view mathematics as facts, rules and formulas in a time when mathematics as a sense-making process is more highly valued in a numerate society" (p. 2). Although the literature revealed that the debates expanded from the US across to other countries, including the UK, a gradual acceptance emerged that learners should develop both their conceptual and procedural knowledge since "pitting skill against understanding creates a false dichotomy… the two are interwoven" (Kilpatrick, Swafford & Findell., 2001, p. 122).

Four years later, Star (2005) suggested that the categorisations of conceptual and procedural knowledge were hindering future research. He proposed reconceptualising procedural knowledge. Using the example of solving linear equations, he compared solutions which slavishly followed a taught procedure with other, much more flexible and potentially more efficient approaches too. For Star, the key issue was flexibility which "is not well explained or even accounted for in typical definitions of conceptual and procedural knowledge" (2005, p. 409). Star introduced the concept of *deep procedural knowledge* whereby learners demonstrated "comprehension, flexibility, and critical judgment and that is distinct from (but possibly related to) knowledge of concepts" (p. 408).

I suggest that the inclusion of critical judgement was crucial, drawing attention to the close relationship between flexibility and efficiency for demonstrating calculation fluency. As noted earlier, Anghileri (2000) and Borthwick and Harcourt-Heath (2007) pointed out that the formal algorithms might not be the most efficient approach for certain calculations. This suggested that learners demonstrating deep procedural knowledge should not merely know how to perform certain calculation strategies, but rather they should also know when not to perform them. Moreover, the literature addressed the impact of exposing learners to multiple strategies; Star and Rittle-Johnson (2008) reported a calculation study with Y7s (N = 132) noting that exposing the learners to multiple strategies led to improved flexibility and problem-solving skills.

Measuring understanding.

The literature consistently highlighted that “Developing measures of the quality of understanding of a given mathematical concept has traditionally been a difficult and resource-intensive process” (Jones, Inglis, Gilmore & Hodgen, 2013, p. 113). These issues tended to surround ensuring the validity and reliability of the chosen measures (Bisson, Gilmore, Inglis & Jones, 2016, p. 141). Based on these concerns from the literature, it was beyond the scope of this PhD thesis to attempt to design a measure for understanding. Hence, my review considered existing measures of understanding which might have usefully been applied to my own study as well as their validity and reliability where available.

Although the above literature highlighted the difficulties of measuring understanding, the lack of available measuring instruments drew further attention to those difficulties. One of the few studies addressing understanding included a measure which was developed in the 1980s. Rather than asking participants to perform calculations using either formal algorithms or alternative strategies, the Chelsea Diagnostic Mathematics Tests (which were originally developed as part of the Conceptions in Secondary Mathematics and Science survey in the 1970s) assessed an individual’s understanding of the four operations by asking participants to identify the correct operation for a particular scenario and “provide ‘short’ stories to fit five given expressions” (Brown, 1987, p. 73). Such an approach enabled a clear focus on understanding but not accuracy or strategy choices, indicating its unsuitability for my own study addressing the four aspects of calculation fluency. Moreover, such an approach appeared extremely time-consuming for a project led by a single researcher.

However, other measures were scarce in the literature. This scarcity necessitated consideration of a series of indicators for learners demonstrating their understanding proposed by Kilpatrick et al. (2001, pp. 118-120). They suggested that if learners fully understood their strategy then they would be less likely to recall it incorrectly, they could identify their errors and make connections in their learning. They also expected learners to possess a comprehensive knowledge of the four calculation operations. Nevertheless, they did not actually trial their indicators with learners and were unable to offer either validity or reliability figures. Rather, they offered a series of categories addressing understanding.

Reflecting on the literature addressing calculation understanding, the lack of measures highlighted the challenges of measuring understanding; hence my decision to adopt the categories proposed by Kilpatrick et al. (2001) for my own study. Although it appeared that the arguments regarding the recognition of the importance of understanding appeared settled, we were still faced with the challenges measuring understanding; especially ensuring the validity and reliability of our measures.

Confidence.

The previous chapter presented confidence as one of the key variables in this study. However, examination of thesauri revealed that *mathematical confidence* and *mathematical anxiety* were potentially synonymous terms. Hence, this section begins by clarifying the definition of mathematical confidence in the literature and exploring the relationship, if any, between mathematical confidence and mathematical anxiety.

Although mathematical confidence and mathematical anxiety might have initially appeared to be closely associated, the literature indicated that they were interpreted in different ways. Mathematical confidence tended to refer to an individual's attitude towards mathematics, addressing their belief in their ability to 'do' mathematics (Fishbein & Azjen, 1975). Hence, an individual's mathematical confidence in this study referred to their general beliefs about their mathematical ability.

Moving on to consider mathematical anxiety, the literature contained a substantial body of research on this topic (including Dowker, Sarkar & Looi 2016; Krinzinger,, 2007; Richardson & Suinn, 1972). This research tended to refer to an individual's emotional state regarding mathematics rather than their attitudes. I suggest that the literature indicated that mathematical anxiety and mathematical confidence should not be regarded as inter-changeable terms. Moreover, "Attitudes to mathematics, even negative attitudes, cannot be equated with mathematics anxiety, as the former are based on motivational and cognitive factors, while anxiety is a specifically emotional factor" (Dowker et al., 2016, p. 4). However, the literature did report a correlation between mathematical anxiety and calculation ability among UK 6- to 9-year-olds (Thomas & Dowker, 2000). Similar findings were also found over a six-month period among 7 to 9-year-old German learners (N = 149) (Krinzinger, Kaufmann & Willes, 2009). My study could explore whether there existed a possible correlation between mathematical confidence and calculation skills.

Having determined that this thesis focused on mathematical confidence, rather than mathematical anxiety, the next step was reviewing the available literature relating to mathematical confidence and calculation. The previous chapter of this thesis noted that UK-based primary-aged boys were reportedly more confident about their mathematical performance than girls (Nunes et al., 2009). The review revealed another UK-based study focusing specifically on confidence and calculation which compared by gender the performance of thousands of 11-year-old learners across a variety of mathematical topics (Johnson, 1987). Johnson's research was based on the large-scale survey conducted by the Assessment Performance Unit (1978-1982) which focused on quantitative data, revealing gender differences in attitudes towards mathematics. The study included both practical and pencil-and-paper tasks. Johnson reported a gender difference favouring boys for practical tasks, such as measurement. For calculation, though, the situation was reversed because the gender difference favoured girls. However, it should be noted that this 30-year-old study pre-dated both the recent reforms and the National Strategies. Johnson also noted that the boys tended to display more enthusiasm and confidence for both the practical and pencil-and-paper tasks than the girls.

Other studies also revealed that more confident learners were more likely to continue to study mathematics post-16 than their less confident counterparts (Brown et al., 2008; Heilbronner, 2013; Pampaka et al., 2011). However, the literature also showed that, even amongst the highest-performing learners, mathematical attainment did not necessarily correlate with high mathematical confidence (ACME, 2012, p. 1).

Noting a decrease in motivation and an interest in studying mathematics as learners progressed through their schooling, Middleton and Spanias (1999, p. 82) highlighted the importance of the learning environment for addressing such concerns. Their findings also raised the possibility of considering school types or teaching styles as factors in my thesis, although incorporating those factors into my own research design might have resulted in an unfeasibly large study for an individual researcher. The next step in this review addresses the different measures for mathematical confidence.

Measuring mathematical confidence.

The literature review revealed several existing measures of mathematical confidence. This section explores the four available measures which had been administered in studies involving under-16s. The following paragraphs explore the similarities and differences between the measures, considers their validity and reliability as well as their suitability for my own study.

Confidence and Doubt Questionnaire.

Dutch researchers Vermeer, Boekaerts and Seegers (2000) reported their findings for their Confidence and Doubt Questionnaire (CDQ) for a calculation and problem-solving study among Y7 learners (N = 158). Their results revealed that the girls were significantly less confident about their problem-solving ability than the boys, but there was no significant difference in confidence for their calculation work. However, the researchers revealed that the CDQ required around forty minutes to administer to each individual Y7 learner, making it a very time-consuming research instrument to consider for my own research design.

Computer-based questionnaire.

Working with 15-year-olds based in Singapore (N = 1,940), Stankov, Lee, Luo and Hogan (2012) collected their mathematical confidence data via a computer terminal. Each participant was asked to rate their confidence on a five-point Likert-scale as they worked their way through a series of mathematical problems on their screens. Although the researchers did not report any gender differences in their findings, they did stress the importance of confidence for predicting mathematical attainment. However, using computers for my own data collection would have presented additional challenges for my research design due to the stringent data protection rules in UK schools as well as the logistical challenges posed by using IT across multiple sites, rendering it almost an almost unworkable approach for a sole researcher. Nevertheless, their use of a Likert-style scale was adaptable for my own study and it also reflected similar data collection methods in the other studies considered in this review.

MaST questionnaire.

A UK government commissioned report noted that mathematics was "a discipline not always embraced with enthusiasm or confidence" (Williams, 2008, p. 1). It led to the establishment of the Mathematics Specialist Teacher (MaST) Programme to train a national network of Masters-level primary teachers to support the learning of future generations of mathematicians. As part of the project's evaluation, Walker et al. (2013) explored the views of the learners of the participating teachers, focusing on their mathematics lessons as well as their mathematical confidence. They reported "a number of positive impacts on learners' attitudes towards, and confidence in, mathematics as a result of the MaST Programme" (p. 130). For the KS2 learners, the first nine questions focused on views about their mathematics lessons and the remaining three questions addressed their mathematical confidence (pp. 135-137):

Learning about maths will help me to get a job when I'm an adult.
I would like to do a job that has some maths in it when I grow up.
I would like to carry on learning maths as I grow up.
I like the way we learn maths.
Is maths one of your favourite subjects?
I like to learn new things in maths.
I find maths interesting.
I enjoy maths.
My teacher helps me understand things in maths lessons.
I do well in maths lessons.
I find maths easy.
I understand maths.

However, Walker et al. (2013) did not disclose details relating to the questionnaire's design, testing or implementation. Due to its apparent lack of rigour, their questionnaire did not appear to be a suitable instrument for my own data collection without considerable further development and testing unless there were no other available measures. Nevertheless, the literature did reveal another potential measure of mathematical confidence for my study.

Competence questionnaire.

Of relevance to my own study, due to its focus on gender and mathematical confidence, was the study by US-based researchers Eccles, Wigfield, Harold and Blumenfeld (1993). They reported that mathematical confidence, which they categorised as *competence,* was higher among 7- to 10-year-old boys than girls (*N* = 865). In their study, the researchers read aloud a series of questions to each class which explored confidence levels in three different academic subjects, including a set of questions specifically addressing mathematics (Figure 7). Reflecting on the wording of these questions, and the extent that they related to this study's definition of mathematical confidence, indicated the relevance of their questionnaire for my own data collection. More specifically, my study adopted the definition of mathematical confidence as an individual's belief in their ability to 'do' mathematics (Fishbein & Azjen, 1975) and the questionnaire's content consistently probed such beliefs by asking questions such as "How well do you expect to do in math this year?" and "How good in math are you?"

How good in math are you? (not at all good, very good)
If you were to list all the students in your class from the worst to the best in math, where would you put yourself? (one of the worst, one of the best)
Some kinds are better in one subject than in another. For example, you might be better in math than in reading. Compared to most of your other school subjects, how good are you in math? (a lot worse in math than in other subjects, a lot better in math than in other subjects)
How well do you expect to do in math this year? (not at all well, very well)
How good would you be at learning something new in math? (not at all good, very good)
In general, how hard is math for you? (not at all hard, very hard)
Some things that you learn in school help you do things better outside of class, that is, they are useful. For example, learning about plants might help you grow a garden. In general, how useful is what you learn in math? (not at all useful, very useful)
For me, being good in math is (not at all important, very important)
In general, I find working on math assignments (very boring, very interesting [fun])
How much do you like doing math? (not at all, very much)

Figure 7. Ten confidence questionnaire items. Reprinted from "Age and gender differences in children's self-and task perceptions during elementary school" by J. Eccles, A. Wigfield, R.D. Harold and D. Blumenfeld, 1993, *Child Development*, *64*(3), p. 834. Copyright 1993 by Society for Research in Child Development. Reprinted with permission.

In a similar approach to several other studies considered in this review, Eccles et al. (1993) adopted a Likert-style response scale in their questionnaire, but also child-friendly images on their scales.

Additionally, Eccles et al. (1993) placed simple labels to the start, middle and end-points to the scale to further support the children completing the questions. Their questionnaire appeared to be highly suitable for my own study due to its extensive piloting with learners of a similar age to my own sample (*N* = 100), followed by its successful administration in their subsequent study which included factor analyses indicating the scale's high internal consistency (0.78) and its validity. The questionnaire could be efficiently administered by teachers to their whole class, using a Likert-style scale and smiling faces, but the terms *math* and *student* would need replacing for my UK-based study.

Gender and Attainment

Terminology.

The literature revealed that the terminology addressing *sex differences* and *gender differences* had evolved over the last forty years. Until the 1970s, the literature showed that researchers tended to refer to sex differences (Figure 8). It was suggested that 'sex differences' implied that those differences were biologically determined and therefore permanent (Fennema, 2000).

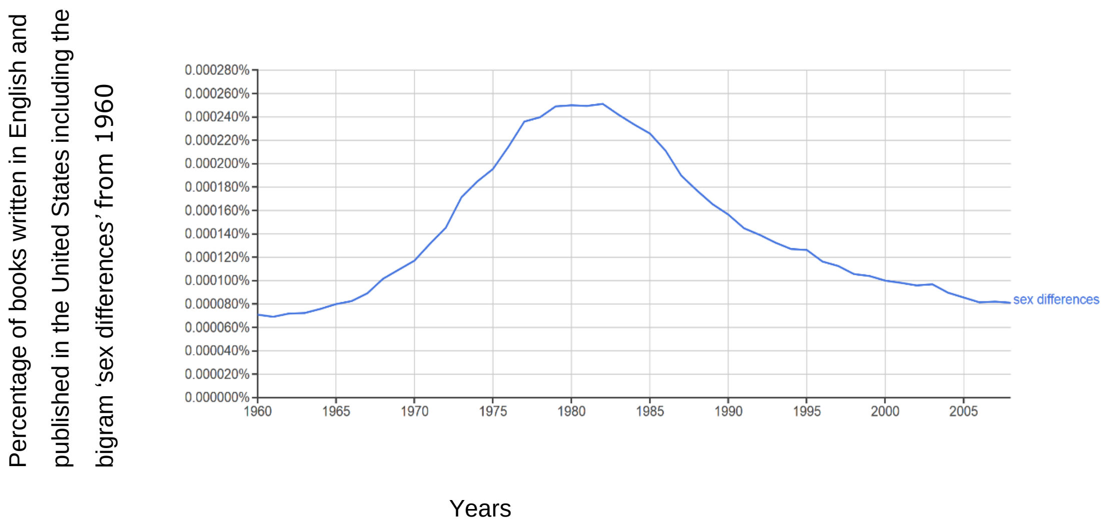


Figure 8. Google ngram viewer results for the percentage of books written in English and published in the United States including the bigram 'sex differences' from 1960 onwards. Accessed from https://books.google.com/ngrams

In the 1980s and 1990s the literature revealed a move towards using the term *sex-related* in research articles (including Caplan, Macpherson & Tobin, 1985; Peterson & Fennema, 1985). In contrast to the term sex differences, sex-related differences appeared to imply an impermanent condition. By the 1990s most researchers were adopting the term gender differences (Figure 9) in their work (including Carr & Davis, 2001; Else-Quest, Hyde & Linn, 2010; Fennema, 2000; Lindberg, Hyde, Petersen &

Linn, 2010). In 2013 the Oxford University Press (OUP) defined gender in its online dictionary as "typically used with reference to social and cultural differences rather than biological ones." Thus, the phrase gender differences was adopted for comparing the performance of girls and boys in this thesis.

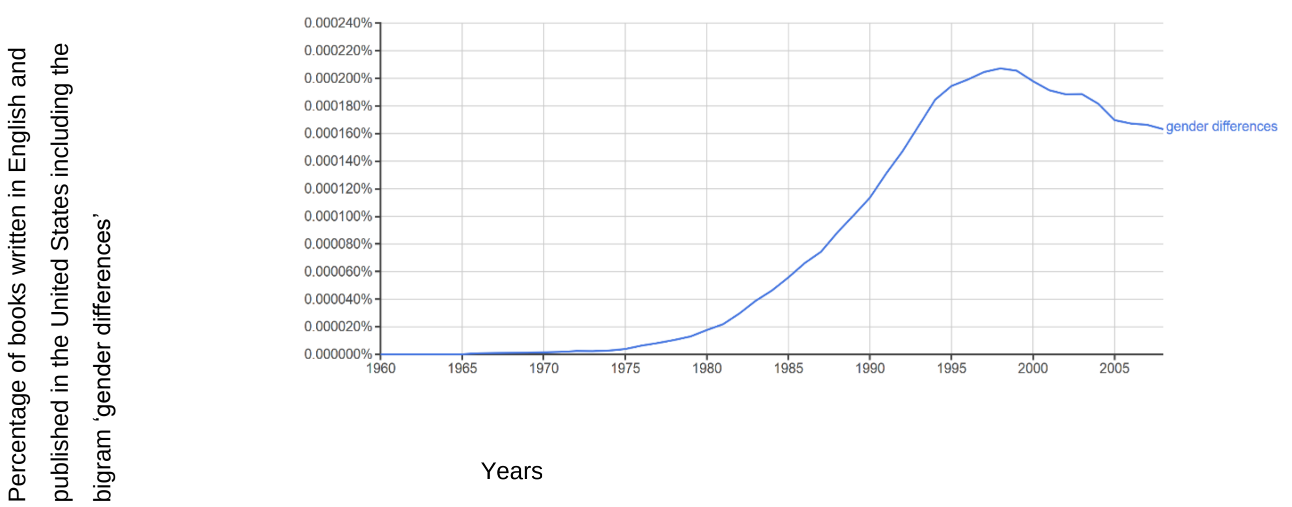


Figure 9. Google ngram viewer results for the percentage of books written in English and published in the United States including the bigram 'gender differences' from 1960 onwards. Accessed from https://books.google.com/ngrams

Gender, attainment and calculation strategies.

Addition and subtraction studies.

Although many under-7s did not use formal algorithms or other written strategies, the available research did reveal gender differences in their approaches towards calculations. In this section I consider US-based research projects focusing on primary-aged learners attempting addition calculations (Bailey et al., 2012) or both addition and subtraction calculations (Carr & Alexeev, 2011; Carr & Davis, 2001 and Fennema, Carpenter, Jacobs, & Levi, 1998). The research design for each of the latter three studies included at least one individual, face-to-face session with each of the participants. During those sessions, the learners were asked to confirm their calculation choices but none of those studies asked the learners to justify those calculation choices.

Looking more closely at those studies revealed an interesting research design adopted by Carr and Davis (2001). Their choice/no-choice approach with 6- to 7-year-old learners (N = 84) allowed them to compare how their participants performed addition and subtraction calculations under different conditions. In the choice activities, learners could have either worked mentally or choose resources, but they were restricted to those two choices and the learners were asked to clarify, but not justify, their strategies with the researcher. I would have liked to discover the reasons behind their calculation choices with the learners themselves. However, a choice/no-choice research design required significant time for its data collection phase, potentially restricting its sample size. In the Carr and Davis (2001) study, during the no-choice scenario, the learners were instructed which approach should be adopted for each calculation. The research design enabled the comparison of learners' preferred strategy choices by gender as well as their ability to calculate using their second-choice strategy. The results showed that the boys tended to prefer mental strategies whereas the girls were more likely to choose resources. When the boys were instructed to use resources, they were as successful as those girls who preferred that strategy. Conversely, those girls who were instructed to use mental strategies were not as successful as the boys. To build on their study, it would be interesting to find out whether participants knew any other calculation strategies because the choice/no-choice research design limited their choices.

A different approach was adopted for a study exploring gender differences for learners moving from single-digit to multi-digit calculations for addition and subtraction (Fennema, Carpenter, Jacobs, & Levi, 1998). In their longitudinal study, the researchers tracked the accuracy rates and strategy choices of 6- to 8-year-old learners. The calculations increased in difficulty as the participants progressed through their schooling, from single-digit calculations in the first year to multi-digit calculations in the third year. After conducting five interviews with each learner during the project, the researchers reported similar findings as Carr and Davis (2001); the boys tended to work mentally and use derived facts whereas the girls were more likely to use counting approaches. They reported no significant gender differences in their accuracy rates, and it would be interesting to explore whether similar findings occurred with slightly older learners in my own study. In the third year of their project, the researchers reported that 8-year-old boys were far more likely to invent strategies and work abstractly whereas the girls tended to choose written strategies. The difference was most notable for subtraction; 80% of the boys invented their own strategies compared to just 45% of the girls. Since my research focused on slightly older learners, it would be interesting to explore whether these gender differences continued or changed as the learners progressed through their schooling for addition and subtraction, as well as explore the situation for multiplication and division.

More recently Bailey et al. (2012) reported gender differences for learners performing single-digit addition calculations in their six-year longitudinal study (N = 229) among 6-to-12-year-old learners. Their research focused on gender differences in the strategy preferences of learners within each year group as well as across year groups as they progressed through their early schooling. Since the researchers were comparing gender approaches with a sample including 6-year-old learners, an age when fewer learners would have adopted written methods, they did not give learners pencil-and-paper. Their findings indicated that the gender preferences identified for addition by both Carr and Davis (2001) and Fennema et al. (1998) continued through the next few years of schooling.

Bailey et al. (2012) also reported a gender difference in attitudes towards accuracy in calculation in their study; the girls appeared to prioritise accuracy, by counting on their fingers, over the efficiency of adopting a mental approach whereas the boys maintained their preference for mental approaches from an early age and their accuracy tended to increase as they grew older. Those findings initially appeared inconsistent when compared with those of Carr and Alexeev (2011) whose longitudinal study tracked over three years from 7-years-old (N = 207). During individual observations, the learners completed age-related addition and subtraction calculations. In contrast to the longitudinal study led by Bailey et al. (2012), but adopting a similar approach, the calculations in Carr and Alexeev's (2011) study ranged from single-digit calculations for the 7-year-old learners to multi-digit calculations two years later. The researchers also adopted a similar research design to the other three studies by asking the learners to confirm their strategy choices with their interviewer, but not the reasons behind those strategy choices. Carr and Alexeev (2011) reported no significant differences for 7- to 9-year old boys choosing mental strategies or using resources. However, they speculated that their choice of multi-digit calculations in their research design might have influenced those boys towards choosing resources since the numbers were more difficult to hold mentally than in the earlier study.

Another recent study, a choice/no choice addition and subtraction study led by Belgian researchers Torbeyns and Verschaffel (2013), reported that the participating 9- and 10-year-old boys (N = 21) tended to prefer the formal algorithm rather than rounding for calculations such as 482 – 299. The researchers did not include any girls in their study. Crucially, the study was conducted after the boys had received just one year's instruction on using formal algorithms, yet they were already becoming the first-choice strategies for the majority of them. It would have been interesting to explore whether girls responded in a similar manner. The researchers suggested that their findings indicated the high prestige afforded to formal algorithms by the boys, reflecting similar findings by Cockcroft (1982) and Fielker (2007). Looking at their research design, Torbeyn and Verschaffel's study adopted the Adaptive Strategy Choice Model (ASCM) first proposed by Lemaire and Siegler (1995, p. 83) which addressed the following four variables regarding an individual's strategy use: which strategies they used; when they used each strategy; how they executed each strategy; and, how they chose each strategy. Lemaire and Siegler proposed that changes in any one of those four variables could yield overall improvements in speed and accuracy. Although their model focused on both speed and accuracy, it did not appear to address the role of the curriculum which the learners were following which was crucial for my own study that followed a change in government policy resulting in the prioritisation of formal algorithms (DfE, 2013a). Moreover, it was recently suggested that the ASCM "may need to be extended to include factors beyond the learner and the problem" (Fagginger Auer, Hickendorff, van Putten & Heiser, 2016, p. 156). If calculation fluency required understanding as well as procedural skills, then my data collection needed to accommodate understanding too.

Division studies.

Although the bulk of the literature involving KS1 learners tended to focus on addition or subtraction calculations, the body of research surrounding slightly older learners was dominated by division studies. For example, several years after the implementation of the National Strategies, Anghileri (2006) reported gender differences in strategy choices for division calculations (*N* = 308); almost half of the boys chose mental or informal strategies for division, compared to just over a quarter of the girls. She concluded that the wider range of calculation strategies encouraged by the National Strategies benefited the 9- and 10-year-old boys because they tended to be more successful than the girls for all of their selected approaches, except chunking down (see Appendix C), compared to her earlier study (Anghileri, Beishuizen, van Putten & Snijders, 1999) which explored division strategies for learners from England and the Netherlands (*N* = 535). Both studies were large-scale surveys using a learner worksheet which was completed during a whole class lesson; this was a data collection instrument which could have been adapted for my own study. Anghileri (2006) reported that the National Strategies were slowly making an impact on schools because learners were choosing a wider variety of division approaches than just a few years earlier. However, the large-scale survey design did not incorporate individual interviews to uncover the reasons behind those calculation choices with the learners themselves.

A more recent Dutch division study of 12-year-olds (*N* = 362) adopted direct observations of learners within a large-scale survey (Hickendorf et al., 2010). They compared individual differences for mental and written strategies, generating their own list of questions to ensure that they posed an appropriate, age-related level of challenge. The overall results revealed that more boys than girls chose a mental approach, in line with the previously discussed research (Anghileri, 2006; Bailey et al. 2012; Carr & Davis, 2001; Fennema et al., 1998). Although Hickendorff et al. (2010) adopted a similar approach to Anghileri (2006), by administering their questions in the classrooms of the 12 participating schools, they also adopted a choice/no-choice research design. In their choice activity, learners could choose whether to make any written notes, which allowed a wider range of responses than the two options seen in the Carr and Davis (2001) study but did not reveal any data regarding the reasons behind those choices. In the no-choice scenario, learners recorded their procedure. A small sample of the participating learners was interviewed who did not record any written notes in the choice task but had recorded answers. Under the no-choice condition, those boys who preferred mental methods in the choice scenario tended to use informal written methods. The written approach was taken by more girls than boys. The results suggested similar findings as the earlier US-based study focusing on addition and subtraction (Fennema et al., 1998): the boys tended to prefer informal written calculation strategies whereas the girls were more likely to adopt formal approaches. Furthermore, when the researchers asked the learners to record their division strategies, the boys tended to record fewer steps and work more intuitively than the majority of the girls in the sample.

The four operations.

Reflecting on the gender difference studies discussed so far, the review revealed a scarcity of studies addressing all of the four operations and addressing this gap in the literature would enable the comparison of strategies both within and between the operations. However, the review did uncover a Dutch study which addressed three out of the four operations exploring the possible connections between mathematical confidence and success at problem-solving and performing calculations (Vermeer et al., 2000). Their study focused on 12-year-old learners (*N* = 178) attempting six age-related calculations and six problems, excluding addition calculations. Each learner also completed the previously discussed confidence questionnaire using child-friendly smiley faces as a Likert-scale to indicate their mathematical confidence. Again, the research design included individual sessions with the participants, but it did not explore the reasons behind their calculation choices. The researchers reported that the girls lacked confidence with their problem-solving questions, but they did not lack confidence with performing calculations.

Formal algorithms and future attainment.

Reflecting on the literature so far, it appears that girls were more likely to choose written strategies over mental strategies for their calculations. The literature addressed the possible longer-term consequences of those confidence issues and strategy preferences; US-based Villalobos (2009) argued that girls tended to benefit more than boys from the focus on number and calculation in the primary curriculum, due to their perceived tendency to prefer following algorithms, but warned that focusing too heavily on algorithms might result in those girls struggling to adapt to choosing different strategies when required to do so by their problem-solving secondary mathematics curriculum. This theoretical model was important for my study because the DfE's reforms (2013a) not only prioritised formal algorithms at primary school, but they also raised the profile of problem-solving in the secondary curriculum. Reflecting on Villalobos' model, her argument offered a possible explanation for the gender difference favouring boys in the mathematical performance of 15-year-old learners since the secondary curriculum was more heavily focused on problem-solving than the primary curriculum which focused on calculation. It appeared that over-emphasising formal algorithms with young learners might have been counter-productive in the longer term for their overall mathematical development. If Villalobos was correct, then the DfE's reforms (2013a) might have inadvertently widened the gender difference in mathematical performance.

Gender activation.

The existing literature also addressed the impact of activating gender stereotypes on mathematical performance. The available studies focused on activating gender stereotypes before completing mathematical tasks and the effects of choosing stereotypical contexts on performance in those tasks. Both aspects revealed crucial considerations for my own research design.

Pre-task gender activation.

The studies which involved activating gender stereotypes before the participants attempted their mathematical tasks were classified as *gender identity* studies. US-based Ambady, Shih, Kim and Pittinsky (2001) explored activating the gender identities of learners from three different age groups (N = 151). To activate the gender identity of their participants, they provided colouring-in activities for their younger learners and questionnaires for the older learners. The researchers reported a significant gender difference in performance among the 11- to 13-year-old learners; the performance of the girls tended to decrease compared to the control group whereas the performance of the boys of the same age was more likely to increase compared to the control group. For the younger learners, the performance of both genders improved after the gender manipulation activity for the 8- to 10-year-old learners. The researchers concluded that the younger learners were more susceptible to a form of chauvinism regarding their gender than the slightly older 11-to-13-year-old learners. Therefore, it was important to consider the possible effect of pre-task gender activation in my research design.

Another gender identity study revealed that classroom organisation could affect the calculation performance of girls (Neuville & Croizet, 2007). The French researchers conducted an experiment exploring stimulating gender stereotypes on the calculation skills of 8- and 9-year-old learners (N = 79). Adopting a similar approach to Ambady et al. (2001), the researchers asked each learner to colour a picture; the girls were randomly assigned either a picture of a girl holding a doll, or a landscape scene, and the boys were randomly given either a picture of a ball or the landscape scene. The colouring-in activity was intended to activate the gender identity of the boys and girls, with the landscape activity acting as a control, before the learners were asked to complete seven calculations. The results indicated that the girls performed better on the less difficult calculations when their gender had been activated by the drawing. However, their performance deteriorated when they attempted the more difficult questions. The results of the boys appeared unaffected by the gender activation. The researchers concluded that classrooms should be organised to avoid gender activation to improve the calculation performance of girls. These findings indicated that my own study needed to consider the possibility of gender activation during the design of my data collection instruments and procedures.

Gender activation during tasks.

The existing evidence base also addressed the effect of question context on gender differences in mathematical performance. An OECD report (2013, p. 82), focusing on learner beliefs in mathematics, noted gender differences in mathematical performance favouring boys when responding to questions which appeared to link to stereotypical situations. For example, a calculation set in a context using petrol received correct answers from two-thirds of the boys but less than half of the girls. Although it was possible to conclude that such issues might have been addressed by substituting contexts deemed boy-friendly for girl-friendly contexts, Zohar and Gershikov (2008) cautioned that such actions might not have necessarily improved the performance of girls. Their study (N = 523) was based around a series of mathematical after-school lessons in Israel. Learners were allocated to one of three contexts; stereotypical boy, stereotypical girl or neutral. The researchers made two key findings pertinent for the design of the calculations for my own thesis. First, the performance of boys was higher for the stereotypical boys' context than the neutral context. Second, the performance of girls varied for the stereotypical girl context according to their age; the younger girls scored more highly than the older ones.

Another US-based study explored the impact of setting *gender neutral* questions on mathematical performance. US-based Che, Wiegert and Threlkeld (2012) set learners the following question*:*

> The capacity of an elevator is either twenty children or fifteen adults. If twelve children are currently in the elevator, how many adults can still get in? (p. 316)

The research team reasoned that their question was gender neutral because it omitted personal names as well as having the equal likelihood that the participating girls and boys had travelled in a lift. In single-sex classes, they compared the problem-solving strategies of 11- to 14-year-old learners (N = 162). Their results indicated that more than half of the girls selected either an additive or procedural approach compared with a quarter of the boys. They also reported that the girls tended to rely more heavily on following algorithms than the boys, echoing the findings of the previously discussed studies exploring gender differences in calculation (Bailey et al., 2012; Fennema et al., 1998 and Hickendorff et al., 2010). Another interesting aspect of their research design for my own study involved their data collection procedures, whereby the questions were completed during a class lesson. During the subsequent analysis stage, the researchers acknowledged their difficulties classifying the calculation approaches of individual learners when a final answer was not recorded.

Reflecting on the impact of the evidence so far regarding question writing for my own study, it was important to set gender neutral questions for the learners to avoid triggering their gender identity. One possible approach might involve setting context-free calculations rather than word problems. To address concerns that some learners might not record their working out, making subsequent analysis of their calculations challenging, the instructions for the learners could encourage them to show their working out. Next in this review I explore the evidence relating to open and closed response types.

Response types.

When given a variety of multiple-choice and open-response questions (N = 154), US-based Gallagher et al. (2000) reported that teenage girls, but not boys, performed better with the multiple-choice questions than the open-response ones. They concluded that, since the girls were more likely to lack confidence in mathematics than the boys, the multiple-choice format perhaps offered greater mathematical support than open-response questions. It was also interesting to consider the evidence relating to response formats and international testing. Exploring the response types for PISA and Trends in International Mathematics and Science Study (TIMSS) revealed a difference in their question formats. Researchers showed that the majority of the TIMSS questions adopted a multiple-choice style format whereas two-thirds of the PISA questions were open-response (Neidorf, Binkley, Gattis & Nohara, 2006). The significant gender difference in England's PISA results, which were not evident in the TIMSS results, raised the possibility of an association between response format and gender performance.

My own research design needed to ensure that my data collection enabled the identification of the calculation strategies chosen by the learners and a multiple-choice format would have limited the opportunities for exploring those calculation preferences whereas an open-response approach would have enabled the learners to record their strategies. Moreover, I needed to consider presenting calculations in such a way that they avoided gender activation both pre-task and during the calculations themselves. Pre-task gender activation could be partly addressed through the procedures shared with teachers, including not indicating the gender aspect of the study with the learners; presenting the learners with open-response, context-free calculations would reduce gender activation during completion of the calculations. The sample questions shared with schools for Y6 SATs adopted such an approach (Figure 10) and I could adapt their question-style for my own study. An additional bonus of this approach was that it meant choosing a data collection approach which was familiar to the learners and their teachers.

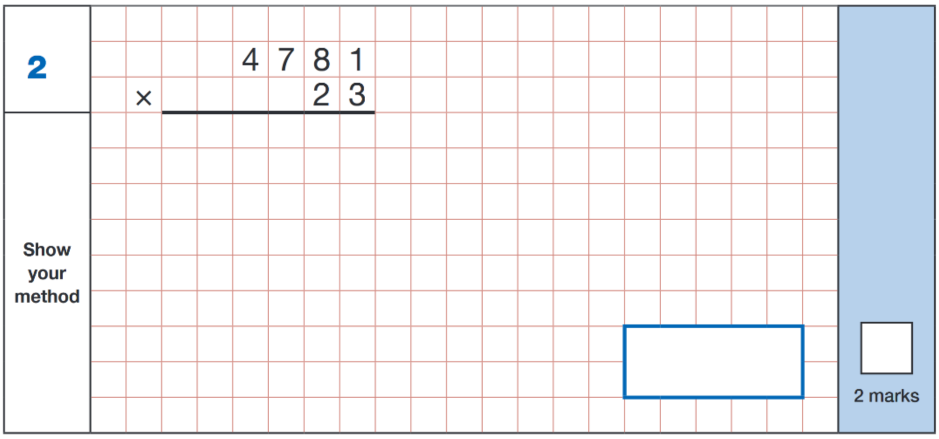


Figure 10. Sample Y6 arithmetic test question. Reprinted from *2017 Key Stage Two Mathematics Paper One: Arithmetic* (p. 11) by Department for Education, 2017.

The Relevance of Written Calculations in the Digital Era

In a world where most adults, and many children, had easy access to either a hand-held calculator or the calculator function on their mobile phone, the continuing presence of written calculations in the school curriculum clearly demanded justification. Although secondary schools had rarely taught the use of log tables or slide rules since they were usurped by calculators in the late 1980s, the teaching of written calculations persisted in the primary curriculum. However, learners required considerable practice before demonstrating their proficiency using either formal algorithms or other written strategies, and some researchers questioned whether that time was well spent (Keiser, 2012; Kilpatrick et al., 2001; Plunkett, 1979). For example, noting that many adults rarely used formal algorithms in their everyday lives, Plunkett (1979) called for schools to rethink their bias towards formal algorithms which caused “frustration, unhappiness and a deteriorating attitude to mathematics” (p. 4). He advocated focusing attention on teaching mental arithmetic, using calculators and developing informal written strategies.

A compelling argument in favour of retaining formal algorithms was their potential for demonstrating proficiency in several mathematical skills. For example, successfully completing the calculation 268 ÷ 13 using the division formal algorithm required knowledge of subtraction skills and an understanding of remainders. Both Ruthven (2009) and Cockcroft (1982) reminded their readers that calculators required specific skills too with Cockcroft noting that:

> a calculator can be of no use until a decision has been made as to the mathematical operation which needs to be carried out and experience shows that children (and also adults) whose mathematical understanding is weak are very often reluctant to make use of a calculator (p. 111).

Moreover, reflecting on the potential of the division formal algorithm for showcasing an individual's calculation skills, Ruthven (2009) called for a "corresponding calculator-mediated procedure to act as a crowning glory" (p. 12).

Others suggested that calculators reduced the need to learn certain written calculations but that they also placed different expectations on future learning. For example, Wolfram (2014) argued that schools should focus on the applications of mathematics rather than continuing to teach strategies such as the division formal algorithm: "We have confused rigour at hand calculations with rigour for the wider problem-solving subject of mathematics" (para. 5).

Despite such calls for increased calculator usage, recent international comparisons revealed a downward trend regarding calculator usage in classrooms (Mullis, Martin & Foy. 2008). The DfE's (2013a) reforms reflected that worldwide trend:

> Calculators should not be used as a substitute for good written and mental arithmetic. They should therefore only be introduced near the end of key stage 2 to support learners' conceptual understanding and exploration of more complex number problems, if written and mental arithmetic are secure (pp. 3-4).

The debate regarding the role of technology was highly relevant for my study. By collecting data on strategy choices of the learners, and reflecting on their efficiency and accuracy rates, it should be possible to review the success rates of their non-calculator strategies and consider whether it might be appropriate to consider encouraging learners to choose using a calculator for calculations where it was deemed more appropriate to do so. However, my study focused on non-calculator strategies; it would also have been interesting to explore the strategies that learners employed when they had the option of using a calculator.

The Anticipated Contribution of This Thesis

It was anticipated that this thesis would contribute towards the literature in several ways. In particular, through its choice of methodology, the development of a new calculation measure and by addressing the relevant gaps in the literature highlighted in this chapter.

Beginning with methodology, the literature highlighted the predominance of quantitative studies for exploring calculation skills which did not reveal the reasons behind the decision-making of the learners. Moreover, these studies tended to adopt a choice/no-choice approach which limited the range of calculation choices for their participants. By designing a study which incorporated both quantitative and qualitative aspects, I anticipated that my thesis would enable the examination of the differing aspects of calculation fluency.

My study would also address the imbalance in the calculation literature whereby calculation studies among younger learners tended to focus on addition and subtraction whereas studies among older learners were more likely to address division. I intended to design a calculation study featuring several calculations for each of the four operations. It was anticipated that my data would enable the comparison of calculation choices across a range of questions so that the study could explore the willingness and ability of learners to vary their strategy choices both within and between the four operations.

It was also anticipated that this study would address several other gaps in the literature discussed in this review. For example, the review indicated a correlation between calculation performance and mathematical anxiety, but it offered no indication as to whether there was a correlation between calculation performance and mathematical confidence, an issue that would be explored in this thesis. Taking into account Villalobos' (2009) model, this thesis would also consider whether there was a gender difference in choosing formal algorithms, and it would use the model to consider the possible implications of my findings for an individual's mathematical development. Finally, my research design would enable addressing gaps in research knowledge relating to calculation efficiency, flexibility and understanding. In particular, my interview data would enable the study of the ways in which learners checked their calculations as well as examine whether they estimated their answers or reflected on their reasonableness of their solutions.

Chapter Three: Methodology

Introduction

The literature review established the existing knowledge base regarding calculation fluency, gender, prior attainment and mathematical confidence. In particular, the review considered the advantages and disadvantages of the differing methodologies and research instruments employed in those studies. This chapter builds on that knowledge in the design of this sequential explanatory mixed methods study. Moreover, it explores my own pragmatic beliefs and how they influenced my design choices. The chapter details the steps taken in the design of the research instruments, including those taken towards ensuring the validity and reliability of the study as well as its approaches regarding the ethical considerations of research too. This is followed by an outline of the steps taken towards the data analysis during the different phases of this mixed methods study.

My Research Philosophy and Epistemological Beliefs

A researcher's theoretical views and assumptions are like a 'fine thread' running through a research design (Burgess, Sieminski & Arthur, 2006, p. 52). Applying the analogy to my own study by acknowledging that my own views and assumptions were woven into my research design, other researchers investigating the same concept of calculation fluency could have proposed alternative research designs due to the influence of their own values and beliefs on the process.

A researcher's values and beliefs have been described using a variety of terms which includes 'worldviews.' According to Guba and Lincoln (1994), worldviews consisted of three components; ontology, epistemology and methodology. Whilst those components allowed for the comparison of different worldviews, Crotty (1998) argued that ontology and epistemology overlapped so that they should be positioned alongside one another. Hence, my own reflections on my worldviews addressed those three components whilst taking into account Crotty's assertion that there is no such clear distinction between their meanings as their separate titles might initially imply to the reader.

Positivism.

A positivist worldview was frequently referred to as 'the scientific method' because it represented the objective stance that physical laws were awaiting human discovery. Consider a physicist who held a positivistic worldview; their realistic ontological stance demanded clearly defined variables for either experimental manipulation or control in order to discover what would have been regarded as scientific fact. Their corresponding epistemological stance implied that, if error and bias had been eliminated, the physicist could uncover the truth. Thus, the value-free foundations of a positivistic worldview

(Kuhn, 2012) clearly contrasted with my previously stated view that beliefs and values underpinned a study. Hence a positivistic approach was rejected for this study.

Post-positivism.

A post-positivist worldview appeared more relevant than a positivist worldview for my research. While post-positivists might have agreed with positivists that a reality did exist, their ontological stance acknowledged that the researcher's own limitations might have restricted their study "imperfectly and problemistically" (Robson, 2002, p. 27). A post-positivist worldview accepted that results were fallible and open to later revision. Unlike a positivist stance, it embraced both qualitative and quantitative data approaches as well as the use of triangulation to reduce errors. However, post-positivism viewed science "as a process of verification" (Hartas, 2010, p. 42), making it incompatible with my own views.

Constructivism.

Crotty (1998, p. 42) defined constructivism as "not believing in the existence of objective truth." Constructivism's ontological stance regarded human knowledge as being constructed, rather than discovered, within that particular moment in time and social context. It led to an epistemological stance where the researcher and subject built knowledge together. I found constructivism a very compelling worldview but accepted that it was incompatible for my study because I was less motivated about finding the absolute truth than enabling teachers to support their learners to develop fluent calculation skills. My ontological stance appeared to fall within another worldview known as pragmatism.

Pragmatism.

Pragmatists regarded knowledge as "an instrument that guides action" (Hartas, 2010, p. 41). Pragmatism arose in the late 1890s from the debates at the US-based Metaphysical Club whose membership included future prominent pragmatists such as Peirce, James, Mead and Dewey. Pragmatists did not consider themselves to be restricted to "any one system of philosophy" (Creswell, 2014, p. 12). Pragmatism offered a worldview aiming to establish 'what works' rather than embracing either subjective or objective epistemological approaches to knowledge (Morgan, 2007, p. 25). The heated paradigm war debates of the 1980s and 1990s featured a "relentless focus on the differences between existing worldviews" (Johnson & Onwuegbuzie, 2004, p. 14). Emerging from those debates, pragmatists were sometimes referred to as 'pacifists' because they were regarded as having brokered peace between different philosophical outlooks (Tashakkoki & Teddlie, 1998). Since my own study was driven by a desire to address the varied aspects of calculation fluency through the most appropriate research design, pragmatism appeared to be a highly relevant worldview.

Popular methodological approaches for pragmatists included action research, design experiments and MMR. However, since both action research and design experiments required a cycle of testing and improving resources which were not relevant for my study, the required combination of quantitative and qualitative data indicated that the most suitable methodology was an MMR design.

A Sequential Explanatory Mixed Methods Research Design

As discussed in the opening chapters, the aim of this thesis was to contribute towards our knowledge regarding calculation fluency, gender, prior attainment and confidence. The research design addressed the four RQs using an MMR design. The following sections outline the nature of that research design in more depth.

Mixed methods research.

This MMR design combined a large-scale quantitative survey with purposefully selected follow-up individual interviews. The large-scale survey focused on calculation accuracy and flexibility, and its findings were explored in more depth during the follow-up qualitative interviews which also addressed calculation efficiency and understanding. Research designs which collected, analysed and integrated both quantitative and qualitative approaches within a single study were generally recognised as MMR. Hence MMR rejected the claims of the incompatibility thesis (Howe, 1988) which asserted that different data types and methods of analysis were incompatible. Indeed, by adopting an MMR design, the possible limitations of a quantitative survey were balanced by the advantages of individual interviews and vice versa. Gorard (2001, p. 4) noted that combining both approaches in a single study was “nearly always more powerful” than restricting a study to an isolated qualitative or quantitative approach. Moreover, neither quantitative nor qualitative approaches alone were sufficient to address my research questions. When combined, quantitative and qualitative approaches offered the potential to complement one another and to provide a more comprehensive study (Greene, Caracelli & Graham, 1989). Moreover, MMR reflected my pragmatic stance which recognised the centrality of my RQs in my study design. Indeed, Teddlie and Tashakkoki (2003, p. 21) suggested that pragmatists regarded their RQs as “more important than either the method they use or the paradigm that underlies that method.” The importance of the RQs was further underlined by Robson (2002, p. 373) who stated that researchers should not be “the prisoner” of their methods or methodologies.

Although the literature review revealed a number of existing calculation studies which conducted large-scale surveys followed by a limited number of individual interviews, their qualitative aspect was designed to confirm rather than explain the initial quantitative findings. The literature also revealed that the number of MMR studies in educational research had steadily increased over the last twenty years, yet my review revealed that it was an unusual for a calculation study; none of the calculation studies unearthed during the literature review adopted an MMR research design. My research design appeared to offer an innovative approach towards investigating calculation fluency. Using an MMR design accommodated researching the differing aspects of calculation fluency in different ways. Moreover, it also addressed two key validity concerns. First, it addressed a type of validity known as ‘paradigmatic mixing legitimation’ (Onwuegbuzie & Johnson, 2006, p. 59). This validity concern cautioned researchers about making tenuous links between their beliefs and MMR designs. In this study, the validity concern was satisfied since my pragmatic beliefs were reflected by an MMR research design which recognised the potential of combining both qualitative and quantitative approaches in order to fully investigate my research questions. Second, by rejecting positivism and embracing an MMR design which demanded switching backwards and forwards between quantitative and qualitative approaches, my research design also addressed the validity concern known as ‘commensurability legitimation’ (Onwuegbuzie & Johnson, 2006, p. 59).

A sequential explanatory design.

Bryman (2006) urged researchers to explain their rationale for choosing an MMR design as well as explicitly sharing their approaches towards integrating the quantitative and qualitative data within their study. Reflecting on his comments resulted in the realisation that the most suitable type of MMR approach for my study was a sequential explanatory design (Figure 11).

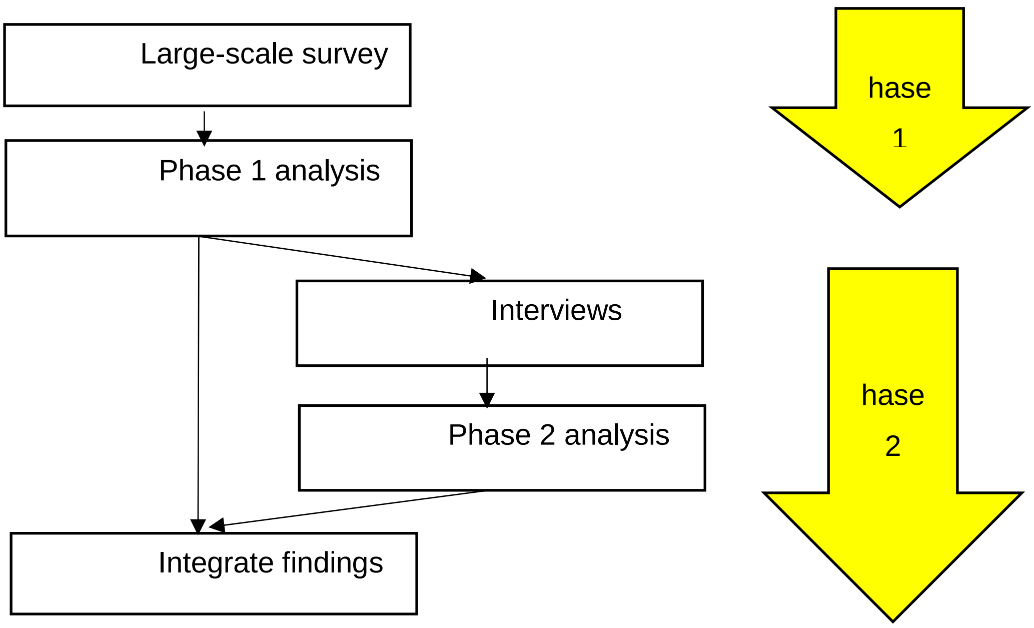


Figure 11. Visual representation of the sequential explanatory research design (Source: Personal collection).

My sequential explanatory MMR design had two distinct phases. The initial, quantitative Phase 1 addressed RQ1, RQ2 and RQ3 by collecting data relating to calculation accuracy and flexibility. The resulting descriptive data informed the selection of participants for subsequent qualitative Phase 2 which was intended to address RQ4. Hence the participants in Phase 2 were drawn from the wider sample in Phase 1, addressing a validity concern known as ‘sample integration’ (Onwuegbuzie & Johnson, 2006, p. 56). Moreover, since the data collected in Phase 2 was also intended to provide explanatory data for the Phase 1 results, the design approach satisfied the definition of a sequential MMR study proposed by Ivankova, Creswell, and Stick (2006, p. 3). They explored the procedural issues surrounding MMR studies, noting both their popularity as well as the challenges that they posed for researchers. Indeed, Bazeley (2002, p. 388) questioned whether some MMR studies were actually “two separate studies which happened to be about the same topic.” Her comments drew attention to the importance of integrating the findings in an MMR study. In my own research, the findings from Phase 1 influenced the sampling for Phase 2. Moreover, the findings from Phase 2 added greater depth to those from Phase 1. In other words, the two phases of the study were closely integrated. Nevertheless, since the qualitative data provided explanatory details for the quantitative data, the quantitative phase dominated the study. Adopting Morse’s (1991, p. 122) symbolism, the sequential nature of the study was represented as QUAN → qual, the capitalisation of its quantitative aspect indicated its dominance, or higher weighting, in the overall research design.

Since the qualitative data collection and analysis was intended to provide a deeper explanation of the quantitative findings, the study satisfied the requirements for an explanatory MMR design. Furthermore, by ensuring that the qualitative phase built on the quantitative phase, the research design addressed specific concerns regarding MMR studies that their different phases could contradict themselves (Robson, 2002, p. 373).

The literature also highlighted several disadvantages of MMR. In particular, as noted earlier by Ivankova et al. (2006, p. 3) MMR required detailed knowledge of both quantitative and qualitative data collection and analysis techniques. It also demanded considerable logistical skills to manage the different phases of the project. Such demands inevitably resulted in an extended data collection and analysis period than might have been required for other research designs. MMR clearly placed a greater burden on the researcher than many other methodologies. Nevertheless, its suitability for addressing my RQs for this calculation study clearly outweighed its acknowledged disadvantages.

The Population

The population for this study consisted of approximately 100 schools enrolled on a regional cohort of the MaST programme. Most, but not all, of the MaSTs were the mathematics subject leader in their school and members of their school's Senior Leadership Team (SLT). As members of their SLTs, the MaSTs were potentially very influential for driving forward innovations in their schools. Thus involving the MaST teachers in both the design and implementation phases of this study offered the potential to address the validity concern known as 'political legitimation' (Onwuegbuzie & Johnson, 2006, p. 59). In other words, my chosen approach was also deliberately intended to enhance the value of this study to the MaSTs.

At this point it should be noted that between 2011 and 2015 I was initially employed as Deputy Director, and later Director, of my regional MaST programme. Before commencing my data collection, I moved onto another post in a different university which involved working on a different project. Since I was no longer involved in the day to day leadership of the programme, I hoped that the MaSTs would not feel obliged to tailor their responses to reflect the teaching approaches on the programme. Also, since the MaST programme has now trained thousands of teachers working in schools across England, I firmly believed that my sample of MaST schools provided as fair as possible a sample of primary schools across the East and East Midlands.

The next three sections will detail how the MMR data collection and analysis was enacted during the study, beginning with the initial quantitative Phase 1. The second and third sections will address the connecting stage and the qualitative Phase 2 respectively.

Phase 1

Cross-sectional survey.

The Phase 1 data was collected via a large-scale survey administered by Y6 class teachers. A survey has been described as the statistical analysis of data collected through questioning (Fowler, 2002, p. 1). The literature review highlighted the benefits of surveys in calculation and confidence research, especially their suitability for collecting data from a large number of participants in an efficient way. Two types of surveys - interviews and questionnaires - dominated the literature (including Creswell, 2012, p. 383 and Hartas, 2010, p. 259). Adopting an interview approach would have demanded considerable resources beyond the means of this study so teachers were recruited to administer questionnaires to their own classes using an approach known as a group-administered survey (Denscombe, 2010, p. 16). This consisted of a calculation workbook and a confidence questionnaire for each learner, with an explanatory script for the teacher to read aloud to the learners during its administration. The validity of the workbook and questionnaire were carefully considered in light of the 'multiple validities legitimation' concern which was regarded as "pertinent to virtually every mixed methods study" (Onwuegbuzie & Johnson, 2006, p. 59). Although the workbook mirrored the survey approach of several existing calculation studies, it was also innovative; the workbook incorporated several examples of each of the four operations, rather than focusing on either just one or two operations or a more limited number of calculations. Meanwhile, the literature review highlighted a confidence questionnaire which had been extensively trialled in other studies and which was adapted for this study. The class teachers who administered the survey also agreed to provide a class list detailing both the gender and KS1 mathematics results for each of the participating learners, the teachers were asked to confirm each learner's gender because the literature revealed that asking learners their gender might have influenced their subsequent responses. Each teacher was provided with a script to read aloud to their learners during the survey, which did not draw attention to the gender aspect of the study, in another effort to reduce possible bias.

In the literature, surveys were typically categorised as longitudinal or cross-sectional. The 'snapshot' nature of a cross-sectional survey was frequently adopted by previous calculation studies. Cross-sectional surveys have been recognised for their straightforward approach to data collection, yet their immediacy also implies that their data could become easily out-dated (Fink & Kosecoff, 1998, p. 52). As a pragmatist, I valued that immediacy since this study was driven by recent changes to the school curriculum (DfE, 2013a).

One of the advantages of adopting a survey approach was that the size of the Phase 1 sample satisfied the demands of statistical testing. Moreover, financial costs were relatively low because the Phase 1 workbooks and questionnaires were not administered by the researcher. Although the literature revealed that a widely acknowledged weakness of surveys was their low response rate (Denscombe, 2010, p. 16 and Gillham, 2007, p. 9), the issue was overcome by asking teachers to administer the survey to their own classes.

Phase 1 sample.

A probability sample was drawn from schools participating in a regional MaST programme. This approach ensured that the study included a manageable group of schools that fulfilled the requirements of statistical testing in the later analysis phases. However, working with group of schools also raises issues regarding clustering since the learners were nested in classes within schools, "Schools, in particular, form natural hierarchical structures providing their pupils with a very specific experience that differs from other schools in the area. This shared experience often translates into subtle differences in outcomes at an aggregated level" (Hodgen et al., 2019, p. 15). Nevertheless, the literature revealed a lack of consensus regarding clustering. For example, Gorard (2007, p. 228) suggested that, "It is perfectly proper to conduct a national survey, and randomly select cases from the same village, street, or even house. This is part of what 'random' means." Researchers often addressed the potential effects of clustering by adopting multi-level models. However, adopting such models "makes analysis more complex, so excluding more readers from its reporting, without clear benefit" (Gorard, 2007, p. 222). This was a crucial point for my research design since my project addressed an issue of national policy and my audience included policy-makers, school leaders and teachers. Although I concluded that the benefits of adopting a single-level model outweighed the disadvantages of a multi-level approach, my decision meant that my findings should be regarded as indicative rather than conclusive.

The Phase 1 sample was estimated to require around 20 schools - its size was influenced by the later data analysis stage. The apparently counter-intuitive importance of the later analysis stage determining the initial sample size was frequently raised in the literature, including Fowler (2002, p. 36) and Gorard (2001, p. 15). It was perhaps most dramatically expressed by Gorard (2001, p. 15) who shared his "quite frightening" realization that an apparently large sample could be undermined during the analysis stage by low numbers of entries into its different categories. This concern was particularly relevant for a calculation study. For example, Borthwick and Harcourt-Heath (2007) recorded up to seven different strategies for each of the four operations, yet some of those strategies were selected by less than ten learners. Guided by their results, and making reference to the relevant statistical tables, my study required approximately 600 students. Since an average primary school had approximately 36 Y6 learners (DfE, 2014a), and taking into account possible withdrawals from the research, 19 schools appeared to be an adequate number for the survey. The sample was drawn from the MaST schools meeting the following criteria:

i. School has Y6 learners.
ii. Y6 teacher willing to support the research, including: sending home consent letters to families and informing myself of any opt-outs; administering the Y6 learner workbooks and questionnaires; returning the completed questionnaires to myself (costs covered by myself); and, if selected, willingness to support the Phase 2 interviews.
iii. School agrees to share participating learners' KS1 SAT results.
iv. School leadership agree to opt-out procedure (this approach is discussed in more depth under the heading Research Ethics); a consent letter would inform parents and carers of potential participants of the aims and tasks involved in the research, only those opting out their child would be removed from the sample.

The eligible schools, which had expressed an interest in the project, were numbered from 001 onwards and random number tables were used to select the actual sample. Phase 1's probability sampling was followed by purposeful sampling for Phase 2, based on the initial findings from the Phase 1 data, meaning that the participants in smaller second sample were drawn from the larger initial sample. The purposeful sample, which is described in more depth on page 82, consisted of groups of learners satisfying criteria identified from the Phase 1 results.

Designing the learner workbook.

The learner workbooks needed to collect data about calculation choices across the four operations as well as a range of calculations. Mindful that the quality of the resulting data depended heavily on the quality of the questions (Taber, 2007, p. 74), the first step was identifying the most appropriate calculations for the workbook. Looking at the calculations chosen for previous studies, they were deemed unsuitable for a variety of reasons. First, the age-groups participating in those studies varied widely, rendering their calculations unsuitable for my Y6 study. Moreover, it was unusual to find any study which addressed all of the four operations - a rare example was the longitudinal study by Borthwick and Harcourt-Heath (2007, 2010, 2012, 2015 & 2016) which included one calculation for each operation of varying levels of difficulty too. Hence the following sub-sections detail how a new set of calculations was designed specifically for this study. Since the calculations had not been validated by a previous study, they underwent a rigorous process of drafting, piloting and expert validation prior to the data collection process.

Each workbook which was based on the style of the Y6 SATs. Throughout my teaching career, colleagues have often presented SAT-style questions to their classes so that their learners became accustomed to their layout. By designing learner workbooks with a familiar layout, it was hoped that the task would appear less intimidating for the learners. Each of the 16 questions was a multi-digit, context free calculation. The decision to include context-free calculations was influenced by the literature review's finding relating to gender activation (Che et al., 2012).

The calculation studies reviewed in preparation for the literature review informed the criteria for the selection and presentation of the calculations for this study as follows:

i. Present calculations horizontally (Bailey et al., 2012)
ii. Randomly order the four operations (Neuville & Croizet, 2007)
iii. Put the smaller number first for at least half the questions (Bailey et al., 2012)
iv. Include calculations requiring ‘carrying’ or ‘borrowing’ for addition, subtraction and multiplication questions (Torbeyns & Verschaffel., 2013)
v. Vary the number of digits in the calculations (Anghileri et al., 1999)

The drafting process.

A draft set of calculations was reviewed by a different cohort of MaSTs than those participating in Phase 1. The review was conducted during a drop-in evening session at a residential event. Around 100 MaSTs commented on the draft calculations. This process offered an initial peer review of the proposed calculations, enabling an 'inside-outside legitimation' review of their validity (Onwuegbuzie & Johnson, 2006, p. 58) prior to the piloting process. Over the course of the evening the MaSTs considered the age-appropriateness of the questions, suggested possible strategies that their learners might choose and estimated the amount of time required complete the calculations. They were also asked to suggest the most suitable time of year for administering the survey to classes and whether or not they should be administered during the daily mathematics lesson.

At this point it should be noted that, due to the MMR nature of the study, the time difference between administering the survey and conducting the interviews needed to be kept as short as possible for two key reasons. First, the short time difference would keep the reasons for the learners' strategy choices fresh in their minds. Second, it reduced the possibility that learners might be taught other strategies between completing their workbooks and attending their interviews. Nevertheless, I also needed sufficient time to review around 600 workbooks and confidence questionnaires in order to support the interview selection process. The MaSTs suggested conducting the survey and interviews either side of a half term holiday, allowing myself a week to review the responses whilst ensuring that the learners received a minimal amount of further mathematical instruction before the interviews. Although schools have three half-term breaks each year, administering the survey in the summer break was immediately dismissed due to end-of-term activities in Y6. The spring break was also ruled out due to its proximity to SATs, the Y6 teachers wanted to avoid disruptions to their classroom routines at that time of year. The majority of the teachers agreed that late October was the most suitable convenient time for conducting the survey, noting that most Y6 learners would be familiar with their calculation strategies by that time. Moreover, the teachers noted that an October survey would also provide the participating schools with valuable information for planning their remaining calculation lessons prior to SATs.

***The draft calculations*.**

The MaSTs supported the identification of a set of 16 draft calculations (Table 4). For each of the four operations, the majority of the calculations were intended to allow learners to demonstrate their ability to select a calculation strategy relating to the features of the calculation (Ruthven, 1995 and Torbeyns & Verschaffel, 2013). For example, although both 501 - 364 and 702 – 695 are both three-digit subtraction calculations, a zero in the tens position or having similar numbers might influence a learner's choice of calculation strategy.

Table 4

Workbook Calculations and Reasons for their Selection

Calculation	Predicted strategy
456 + 372	Formal algorithm
203 + 401	Mental
245+ 246[a]	Near doubles
299 + 534	Round
632 - 154	Formal algorithm
382 - 199	Round
702 - 695	Count-up
500 - 76	Number bonds or count-up
27 x 63	Formal algorithm or grid[b]
35 x 99	Round
20 x 46	Double
568 x 34	Formal algorithm or grid
517 ÷ 19	Formal algorithm
480 ÷ 20	Multiples of 10 or count-up
693 ÷ 3	Divisibility rules
401 ÷ 25	Count-up

Note. [a]calculation changed for full study to 245 + 256, [b]several schools taught the grid method as their main written multiplication strategy

Phase 1 pilot study.

The importance of piloting the survey materials was stressed in the literature (including Cohen, Manion & Morrison, 2007 and Oppenheim, 1992). Indeed, Oppenheim (1992, p. 48) recommended that all of the pilot materials should be tested "even the typeface or the quality of the paper." The following paragraphs demonstrate how those recommendations were followed in this study.

The pilot study was conducted in a primary school in the same geographical area as the proposed full study. Moreover, since the school had not participated in the MaST programme, the piloting process enabled further 'inside-outside legitimation' of the materials (Onwuegbuzie & Johnson, 2006, p. 58). The head teacher agreed to an opt-out approach to parental consent whereby the parents of the

23 Y6 learners (12 boys, 11 girls) were sent the draft opt-out letter, which was printed on school-headed notepaper, a week before the pilot study. During the pilot study, there were no parental withdrawals from the process and the class teacher did not withdraw any learners for either academic or behavioural reasons.

The pilot consent forms, learner workbook, learner confidence questionnaire, teacher script and interview protocol were all trialled in the pilot school. Since the pilot study was conducted with a single Y6 class, the Phase 1 surveys were administered one weekday morning and the Phase 2 interviews were held the following afternoon. The findings from the piloting process relating to Phase 1, which informed the further development of the survey materials, are outlined in the next two sections. The findings from the piloting of the Phase 2 materials will be addressed later in this chapter.

Phase 1 of the pilot was administered by the class teacher, but it was also observed by the researcher in order to support the further development of the materials. During the piloting process, the learners were sat in their usual classroom seats. To begin the session, the class teacher read aloud the pilot teacher protocol which included an instruction that learners should use their preferred strategies, so the researcher could learn more about the ways they preferred to calculate. However, after one of the learners asked if they should therefore choose a different strategy each time, it was decided with the teacher that the sentence should be removed from the final version of the teacher protocol. Other minor modifications were made to the protocol based on the feedback from the teacher. In particular, the teacher suggested stating that the learners should work independently and wait until instructed to start by their teacher.

The piloting process highlighted three key issues regarding the organisation of the workbook session. First, classrooms displays would be normally covered over during SATs but during the pilot process some of the learners were looking at their classroom's calculation displays. Hence, teachers administering the workbooks in the research were asked to cover mathematical displays prior to handing out the workbooks. Second, although most of the learners completed their workbook within the 35-minute timeframe recommended during the drafting process, several of the learners had not finished the task at that point. For the full study teachers were asked to allow the learners slightly longer than 40 minutes to complete the workbook. Thirdly, they were also asked to adopt the SAT approach by seating the learners at individual desks during the survey. The advice given to teachers is presented in Appendix D.

Part of the piloting process involved trialling an initial coding scheme which was based on the list of calculation strategies recorded by Borthwick and Harcourt-Heath (2007) (Table 5). However, they devised a coding scheme for each category which I attempted to simplify by applying it across all four categories (Table 6). Trialling my initial coding scheme with a set of workbooks indicated a number of further revisions for consideration before attempting to code the calculations from the full study. In particular, the piloting process highlighted the challenge of differentiating between counting-up, chunking up and using a number line. The decision was made to combine those three strategies into a single ‘counting-up’ category. Likewise, counting back and using a number line were combined into a single ‘counting back’ strategy. However, since it had a very clear layout, ‘chunking down’ was kept as a separate category. The negative number strategy, though, was not chosen by any of the learners in the pilot study so it was dropped. The ‘two partial products’ category was also dropped because it was deemed to highlight an error rather than a multiplication strategy; those answers were coded under the ‘partition’ category.

Table 5

Coding Scheme by Borthwick and Harcourt-Heath (2007)

	Operation			
Strategy	Addition	Subtraction	Multiplication	Division
Formal algorithm	✓	✓	✓	✓
Number line	✓	✓		✓
Partition	✓			
Expanded vertical	✓		✓	
Count-up	✓	✓		
Count back		✓		
Negative number		✓		
Grid			✓	
Two partial products			✓	
Chunk up				✓
Chunk down				✓
Answer only	✓	✓	✓	✓
Not attempted	✓	✓	✓	✓
Other	✓	✓	✓	✓

Note. Adapted from Calculating: What can Year 5 children do? A. Borthwick and M. Harcourt-Heath, 2007. Copyright by Alison Borthwick and Micky Harcourt-Heath. Adapted with permission.

Table 6

My Initial Coding Scheme

Calculation strategy	Code
Formal algorithm	A
Not attempted	B
Answer only	C
Number line	D
Partition	E
Expanded	F
Negative number	G

Count-up	H
Count back	I
Grid	K
Two partial products	L
Chunk down	M
Chunk-up	N
Other	O

Note. Adapted from *Calculating: What can Year 5 children do?* A. Borthwick and M. Harcourt-Heath, 2007. Copyright by Alison Borthwick and Micky Harcourt-Heath. Adapted with permission.

The 'not attempted' category was changed too, if a learner either did not attempt or complete a calculation, then their response was coded as 'no answer.' Moreover, having reviewed the learner's strategies in the pilot study, the updated coding scheme also included 'doubling,' 'rounding' and 'multiples of ten' (Table 7). The coding process was supported by examples of the different strategies (Appendix C).

Table 7

My Final Coding Scheme

Calculation strategies	Code
Formal algorithm	A
Not answer	B
Count-up	C
Count back	D
Partition	E
Expanded	F
Grid	G
Chunk down	H
Answer only	K
Double	L
Round	M
Multiples of ten	N
Number bonds	O
Other	P

Note. Adapted from *Calculating: What can Year 5 children do?* A. Borthwick and M. Harcourt-Heath, 2007. Copyright Alison Borthwick and Micky Harcourt-Heath. Adapted with permission.

After piloting the workbooks, the next step involved further validation of the calculations. Following the advice of Gehlbach and Brinkworth (2011), the validation process aimed to establish the relevance of each calculation as well as consider any other calculations which had been overlooked in the research design. The validation panel consisted of two 'experts' and three 'lay people,' fulfilling the panel specification recommended by Rubio, Berg-Weger, Tebb, Lee and Rauch (2003). The two experts were Senior Lecturers in Education (Primary Mathematics) whose main role involved working with trainee primary school teachers. The three lay people were members of their primary school's

SLT. This approach offered a third peer review of the calculations during their development, continuing the 'inside-outside' approach taken towards validity in this study (Onwuegbuzie & Johnson, 2006, p. 58). Each member of the review panel was asked to complete a questionnaire which asked them to rate the appropriateness of each calculation in respect of the research questions, comment on the clarity of the workbook and identify any aspects that they felt had been overlooked in the research design (Appendix E). The calculations that they were shown were not placed in order of operation or perceived difficulty, except that the first calculation adopted the commonly-adopted approach of including a straightforward addition calculation to help settle the learners into the task. Then the inter-rater agreement (IRA), ranging from zero to one, was calculated for those aspects (Table 8).

Table 8

Validation Panel Feedback

	Age-appropriateness of calculation							
	Expert							
Calculation	1	2	3	4	5	*M*	*SD*	IRA
456 + 372	4	4	4	4	4	4.00	0.00	1.00
203 + 401	4	4	4	1	2	3.00	1.41	0.60
245 + 256	4	4	4	2	4	3.60	0.89	0.80
299 + 534	4	4	4	2	4	3.60	0.89	0.80
632 - 154	4	4	4	4	4	4.00	0.00	1.00
382 - 199	4	4	4	2	4	3.60	0.89	0.80
702 - 695	4	4	4	2	2	3.20	1.10	0.60
500 - 76	4	4	4	4	4	4.00	0.00	1.00
27 x 63	4	4	4	4	4	4.00	0.00	1.00
35 x 99	4	4	4	4	4	4.00	0.00	1.00
20 x 46	4	4	4	4	2	3.60	0.89	0.80
568 x 34	4	4	4	4	4	4.00	0.00	1.00
517 ÷ 19	4	4	4	4	4	4.00	0.00	1.00
480 ÷ 20	4	4	4	4	4	4.00	0.00	1.00
693 ÷ 3	4	4	4	4	1	3.60	0.89	0.80
401 ÷ 25	4	4	4	4	2	3.60	0.89	0.80
M	4.00	4.00	4.00	3.31	3.31			
SD	0.00	0.00	0.00	1.08	1.08			

Note. IRA = inter-rater agreement.

[a] minimum score = 1, maximum score = 4

It should be noted that an IRA score of at least 0.8 was recommended for new measures (Rubio et al, 2003). The IRA score for the clarity of the workbook was 1.00 and the overall age-appropriateness of the calculations was rated 0.86, both scores satisfying the requirements for new measures. The panel's feedback regarding the clarity of the workbook included further modifications to mirror the SATs even more closely by increasing the amount of colour, changing the font of the writing and adding a pencil graphic to indicate where learners should write their answers. Each of those suggestions was incorporated into the revised booklet. Meanwhile, both of the calculations which received the lowest IRA scores were reviewed and one of them, 203 + 401, was replaced by the calculation 5412 + 2584. Although the substitute calculation could also be performed mentally, since it did not require any ‘carrying,’ it was not as obvious with a four-digit number. However, the calculation 702 – 695 was retained despite its similarly low IRA rating. Although it appeared to be a simple difference to calculate mentally, I believed that it posed a challenge for learners using the formal algorithm due to the position of the zero. Hence the decision was made to retain the calculation. The remaining feedback comments from the members commended both the non-linear progression of the calculations and operations as well as the variety of possible calculation approaches for the selected calculations. The final version of the workbook is shown in Appendix F.

Designing the confidence questionnaire.

Although each of the four confidence questionnaires investigated during the literature review adopted Likert-style scale, the differences between them were important to consider for my own research design. The existing questionnaires varied in the way in which they were administered to the learners. They also varied in their use of images, such as stars or smiling faces, on their Likert-style scales as well as varying in their labelling of those scales too. The version of the questionnaire devised by Eccles et al. (1993) was chosen for this study since it had been extensively trialled with primary-aged learners. However, although it was written in English, its research design included various terms such ‘math’ and ‘student’ for US-based schools. Those terms were replaced with ‘maths’ and ‘learner’ for this UK-based study. Adopting a similar approach to Eccles et al. (1993), the class teachers in my study read aloud the ten questions to their learners. Moreover, the Likert-style scales were illustrated with child-friendly faces and labels were added too.

Piloting the questionnaire.

The trial of the pilot confidence questionnaire began by the class teacher reading aloud each question and allowing the learners time to record their answer before moving on to the next question, following the approach adopted by Eccles et al. (1993). The confidence questionnaire was administered before the workbooks were handed out. The teacher was asked to feedback on its appropriateness for Y6 learners, especially the clarity of its language and anticipated ease of completion. The teacher suggested replacing the five smiling faces for each question, noting that a set of faces which gradually changed in five steps from an unhappy face to a smiling face should make the scale even clearer for young learners. This change took place before the full study. The teacher also confirmed that the language in the questionnaire appeared suitable for Y6 learners and that the questionnaire was straightforward to complete too. The pilot questionnaire took around ten minutes to complete in the classroom. Therefore, both the questionnaire and workbook were completed in less than an hour during the pilot study. The final version of the confidence questionnaire is presented in Appendix G. Its reliability was confirmed by calculating its Cronbach's Alpha score which is described in more detail under the Confidence heading in Chapter Five.

Designing the teacher scripts.

The literature review highlighted the possibility of external factors affecting the results of a study, such as gender activation (Ambady et al. 2001 and Neuville & Croizet, 2007). Therefore, in an effort to reduce that effect as well as ensure a consistent approach across the schools, each teacher was asked to read aloud a prepared script which did not refer to the gender aspect of the study (Appendix H).

Phase 1 Data Collection.

As discussed earlier, the survey was conducted prior to the October half-term break. The survey materials were sent out to the participating schools two weeks before the break and each school was contacted to confirm that they had received their survey packs and understood the survey arrangements. All of the surveys were collected by couriers before the schools broke up for their holiday. By the time the schools returned after their half-term break, the descriptive analysis had already been completed and a summary was returned to each school.

Connecting the Phases

In this sequential explanatory mixed methods study, one of the roles of the qualitative Phase 2 was providing a deeper explanation of the data collected in the quantitative Phase 1. This approach offered the potential to increase the validity of the findings in three ways. First, since the interviews were designed to add depth to the survey findings as well as address the outstanding aspects of calculation fluency from Phase One, the research design addressed the validity concern known as 'weakness minimization legitimation' (Onwuegbuzie & Johnson, 2006, p. 58). Second, the data collected in Phase 2 was designed to complement the data collected in Phase 1. This addressed the validity concern, known as 'sequential legitimation' (Onwuegbuzie & Johnson, 2006, p. 58), which challenged researchers to consider the potential impact of reversing the phases in their studies; if the two phases in this study had been reversed, then the number of learner interviews required to provide sufficient data to address the RQs would have increased beyond manageable levels for an individual researcher.

Third, by interviewing a sub-sample of the Phase 1 participants, the research design increased the possibility of yielding quality meta-inferences from the resulting data.

Analysis

Each learner was assigned a unique identification code (UIC). This code was written on the front of their named workbooks and questionnaires. However, only their UIC was entered on to the computer spreadsheet. This approach ensured that none of the real names of the learners were stored electronically during this study. Individual learners were identified for Phase 2 by using their UIC from the spreadsheet to identify their workbook and hence their name. This approach safeguarded their anonymity during the data analysis process and satisfied British Educational Research Association (BERA, 2011) guidelines. The total scores for their confidence questionnaires and workbooks were calculated and entered into the Excel spreadsheet. Then, the calculations in the workbooks were coded in readiness for both the statistical analyses of the Phase 1 data.

Phase 2

Hierarchical focusing interviews.

The literature revealed a wide variety of possible structures and approaches for the Phase 2 interviews. They ranged from highly structured interviews to unstructured interviews, and semi-structured approaches were positioned on a continuum between those other two approaches (Robson, 2002, p. 270). For this study, the interviews had two distinct functions. First, they needed to elicit the reasons behind the calculation choices of the learners. Second, they were required to complete the data collection for calculation fluency. Taking into account those requirements, a semi-structured interview emerged as the most suitable approach because the format combined predetermined questions with the flexibility of modifying other questions based on the responses of the interviewee (Robson, 2002, p. 270). The semi-structured interviews adopted the ‘hierarchical focusing’ approach proposed by Tomlinson (1989, p. 155) who summarised the approach as “having it both ways,” encouraging researchers to maximise both verbal and non-verbal prompts to “elicit as spontaneous a coverage of the interview agenda as possible.”

The interview format consisted of an initial question, intended to encourage learners to share the reasons for their calculation choices, followed by a series of closed, follow-up questions covering any of the outstanding aspects of calculation fluency which the learners did not volunteer during their initial answer (Appendix I). The explanatory nature of the study demanded an interview schedule which explored more deeply the accuracy and flexibility aspects of calculation fluency from Phase 1 as well as the two outstanding aspects of efficiency and conceptual understanding. Hierarchical focusing offered an opportunity to address the various aspects of calculation fluency raised in the literature review whilst minimising my influence on the learners' responses, satisfying the validity dilemma regarding the relative roles of the interviewer and interviewee (Tomlinson, 1989, p. 155). The impact of the interviewer on the validity of their data was frequently raised in the literature (including Aldridge & Levine, 2001, p. 53, Cohen et al., 2007, p. 150; Fowler, 2002, p. 136 and Hartas, 2010, p. 228). Indeed, Fowler (2002, p. 136) suggested that the interviewer deserved a "central place" in the research design whilst Loosveldt (2008, p. 214) suggested that interviewers could be responsible for both random and systematic errors. Potential random errors, such as mistakes recording the data or placing unintentional emphasis on certain words during the interview process, were addressed by taking two recorders to every interview (in case one of the recorders failed) as well as rehearsing and refining my interview procedures. Addressing systematic errors was regarded as the most practical approach towards improving the validity of an interview (Cohen et al., 2007, p.150) and a key concern was the possibility of bias due to failing to ensure a high level of consistency between interviews (Oppenheim, 1992, p. 147). For example, questions could be framed in different ways with different learners, or the questions might have been interpreted inconsistently by interviewees. Hence, trialling with friends and colleagues before the piloting process enabled me to develop as consistent an approach as possible. It was hoped that the hierarchical focusing approach itself would reduce bias because its design was intended to allow each learner's own reasons for choosing calculation strategies to emerge during the open question section of their interviews. Nonetheless, a note of caution was proffered by Taber (2007, p. 115) who acknowledged the benefits of a degree of flexibility during interviews, such as clarifying a question for an interviewee. Also, the learners could inadvertently affect the validity of their responses by aiming to appear in a favourable light, a validity concern known as 'acquiescence bias' (Hartas, 2010, p. 228). Therefore, they were reminded at the beginning of their interviews that the focus was on the interviewer learning their views.

In the available literature, studies employing hierarchical focusing interviews tended to involve either adults or older teenagers (including Taber, 2007 and Tomlinson, 1989). Prior to my own study, it appeared that the approach had not been widely deployed with younger learners. Aware that my interviews required a sensitive environment, as well as needing to avoid the acquiescence bias referred to in the previous paragraph, at the start of each interview I stressed that I was keen to learn more about their ideas. Moreover, to establish a friendly and welcoming atmosphere, each learner was informed at the beginning of their interview that they would receive a 'goody bag' for their help. Their marked workbooks provided a useful focus for the learners and if any interviewee gave an unclear answer, I clarified their response before moving on to the next question.

Phase 2 sample.

Due to the explanatory nature of this study**,** the potential interviewees were selected from the original quantitative sample (*N* = 590). This approach satisfied the definition of a purposive sample which occurs when ‘‘particular settings, persons, or events are deliberately selected for the important information they can provide that cannot be gotten as well from other choices’’ (Maxwell, 1997, p. 87). Moreover, it followed an approach adopted in many other sequential studies where a purposive qualitative sample was a subsample of an earlier quantitative sample (Teddlie & Yu, 2007). The Phase 1 analysis, which is presented in more depth in the next chapter, informed the selection of the interviewees. More specifically, the analysis showed that gender, high attainment at KS1 and low confidence were significant predictors of use of the formal algorithm whilst high attainment and low confidence, but not gender, were significant predictors of calculation accuracy. Three of the Phase 1 schools were randomly selected for Phase 2 using random number tables. Using the analysis from Phase 1, the selection process identified two groups of learners. In the first group were the higher attaining learners who achieved Level 3 at KS1. The second group consisted of learners whose confidence scores were in the lowest quartile for their school. The two groups were not intended to be exclusive since it was possible for the confidence level of a Level 3 learner to fall within the lower quartile. However, none of the learners from the three schools satisfied both of the criteria. For those two groups of learners, their accuracy and use of the formal algorithm scores were ranked and those with the median scores were invited to attend an interview. My approach addressed concerns that any outliers, such as exceptionally able learners, might distort the data. Altogether, 24 learners were invited for interview (12 boys, 12 girls). However, since one of the selected schools had insufficient learners to satisfy the selection criteria, it was replaced by another randomly selected school from the Phase 1 sample.

Phase 2 pilot study.

For the pilot study, eight learners were randomly selected for interview. The interviews were conducted in the corridor outside their classroom. At the start of each interview, the learner was reminded about the overall purpose of the study and they were asked to confirm their willingness to be interviewed. The pilot interviews were recorded on two iPhones. Although the recordings were fairly clear and relatively straightforward to transcribe, some words were indistinct. Moreover, the learners were easily distracted by the iPhones. Therefore specialist voice recorders were acquired for the full study which improved the quality of the recordings, making transcription much easier, and they were much less distracting for the learners too.

The pilot study interviews lasted between 12 and 20 minutes. On reflection, I realised that I had failed to “listen more than you speak” (Robson, 2002, p. 274). I had inadvertently reverted to my teaching role, trying to address their errors and develop their understanding rather than make good use of prompting to encourage the learners to share their own thoughts. In the actual Phase 2 interviews, I focused more heavily on using their workbook responses as prompts to encourage the learners to talk more freely and I made much greater use of a range of non-verbal cues, which was an approach I adopted after re-reading Tomlinson (1989). Those revisions resulted in more detailed responses in the full study than the pilot.

In the pilot study, the learners responded positively to the hierarchical approach. They had completed the calculations the previous day so they easily recalled the reasoning behind their strategy choices, reinforcing the importance of ensuring a short time difference between completing the workbooks and conducting the interviews in the full study. However, there was a minor concern regarding those learners who had chosen the same strategy for each of the four calculations for a particular operation because their line of questioning appeared much more repetitive than for the other learners who had varied their approaches. Furthermore, during the design phase of the workbooks, the feedback indicated that the order of the calculations should vary between the four operations. However, the pilot interviews revealed an unforeseen consequence of that approach. As I flicked through the workbooks searching for the next calculation for a particular operation, some of the learners became distracted by their answers. On reflection, and after discussion with colleagues, I realised that it would be more efficient to photocopy each workbook and reorganise the answers by operation before the Phase 2 interviews during the interview, which proved much less distracting for both the interviewer and interviewee.

The pilot interviews also contributed towards addressing a specific validity concern known as ‘the interviewer effect’ (Denscombe, 2010, p. 193). More specifically, the class teacher listened back to the recording of the interviews and confirmed that their individual learners had given expected responses to the questions. The final version of the interview protocol from the full study is presented in Appendix I.

Analysis, including integrating the phases.

The research design influenced the approach taken towards the qualitative data analysis. A grounded analysis approach was rejected due to the need to address several questions arising from the initial quantitative phase, and thematic or framework analysis appeared to be the most suitable options for analysing the interview data. The final decision was informed by two key issues. First, unlike thematic analysis, framework analysis accommodated *a priori* issues as well as emerging ideas during the analysis process which indicated its suitability for a sequential, explanatory study. Moreover, the perceived transparency of the framework analysis approach for analysing qualitative data was regarded as highly suitable for studies addressing topics relating to government policies (Srivastava & Thomson, 2009). This was an important consideration for my study because it was motivated by changes to the school curriculum (DfE, 2013a).

Framework analysis arose from the research of Richie and Spencer (1994) who developed the following five steps for conducting framework analysis: familiarisation with the data; identifying a thematic framework; indexing the data; charting the data; and mapping and interpreting the data. Their work focused heavily on applied policy research and they recognised the importance of visibility in their qualitative analysis. By developing such an accessible approach towards their qualitative analysis, they attempted to ensure that policymakers would understand how they reached their findings. This was an important aspect for my own study which was driven by recent reforms to the primary mathematics curriculum (DfE, 2013a). Nevertheless, Cohen et al. (2007, p. 469) cautioned researchers that qualitative analysis was highly dependent on the researchers themselves, stressing that they should maintain a high degree of self-awareness throughout the process. On the other hand, another advantage of adopting the framework analysis approach was the frequent opportunities to revise and refine the thematic framework and subsequent charts during the process. Moreover, framework analysis allowed straightforward scrutiny of the evidence for individual learners as well as enabling comparisons between learners using the charts developed during the process. Hence the process could focus on the views of the learners whilst reflecting the design aims of this sequential explanatory study. Indeed, as Srivastava and Thomson (2009, p. 72) noted, “Framework analysis provides an excellent tool to assess policies and procedures from the very people that they affect.”

The process began with the familiarisation stage which involved reading and re-reading the transcribed interview notes in order to become totally immersed in the data. The *a priori* themes were calculation accuracy, use of the formal algorithm (which reflected the flexibility aspect of calculation fluency), conceptual understanding, efficiency and the most accurate calculation strategies. The notes taken during the familiarisation stage were organised into a draft thematic framework which was trialled with a limited number of the transcripts. Richie and Spencer (1994) noted that initial thematic frameworks tend to focus more heavily on *a priori* issues whereas later revisions would accommodate any themes and sub-themes emerging from the deepening analysis. Indeed, my draft framework became much more detailed through the addition of emerging sub-themes following a cycle of trialling, refining and re-trialling my ideas. At first, they were tested on the limited number of the transcripts. Once it was felt that the number of sub-themes had stabilised, the framework was further tested on two of the other transcripts too. The process demanded making judgments about the relevance and importance of the various issues arising from the data in order to achieve a manageable framework which addressed the key aims of the study. The revised framework retained its original themes. The interview transcripts were indexed according to the finalised thematic framework in readiness for the charting process. During indexing, each transcript was annotated to highlight the various themes and sub-themes within it. Charting the data moved the analysis from working with individual transcripts towards analysing the whole data set. Each chart represented a single theme and the data was organised into the sub-themes identified in the previous stages of the process. The completed charts allowed the data to be analysed both vertically (by sub-theme) and horizontally (by individual learner). The charts enabled counting of the total number of coded items as well as consideration of the actual text. This approach addressed the validity concern known as ‘conversion legitimation’ whereby inaccurate weighting could be assigned to different aspects of the data (Onwuegbuzie & Johnson, 2006, p. 58).

Validity and Reliability

A quality research design should be trustworthy, a characteristic primarily judged on its approach towards ensuring a valid and reliable design (Robson, 2002, p. 93). Hence the following two sections describe the approaches taken to ensuring my research design addressed both of those criteria, beginning with validity.

Validity.

The validity of a study refers to the appropriateness of its data for addressing its research questions (Denscombe, 2010, p. 298). Hence the validity of my study was judged according to the appropriateness of the data collected from the calculation workbook, confidence questionnaire and face-to-face interviews. The importance of validity was highlighted by Hartas (2010, p. 74) who argued that it was at the “heart of the enquiry.” Moreover, since Onwuegbuzie and Johnson (2006) suggested that mixed methods studies presented a unique range of validity or ‘legitimation’ concerns, steps were taken to address each of their nine validity concerns throughout the development of the research instruments and the subsequent data collection. Those steps were highlighted at the relevant points in each chapter and a summary is presented in Table 9.

Table 9

The Nine Validity Concerns Proposed by Onwuegbuzie & Johnson (2006) and How They Were Addressed in the Research Design

Validity concern	Steps taken to address validity concern in the research design
Sample integration	Interviewees for the qualitative phase were drawn from the quantitative phase
Inside-outside	Proposed calculations were peer reviewed
Weakness minimization	Interviews provided depth to survey results
Sequential	Phase 2 data complemented Phase 1 data
Conversion	Qualitative charts enabled counting of the coded items as well as consideration of the actual text.
Paradigmatic mixing	Research design reflected the pragmatic approach towards the study
Commensurability	Research design required switching backwards and forwards between quantitative and qualitative data
Multiple validities	Specific validity concerns for quantitative data were addressed during the research design. A similar approach was taken for the qualitative data collection too.
Political	MaSTs involved in both study design and its implementation

As well as the validity concerns arising from the selection of an MMR design, the validity of the research instruments was also taken into consideration. The learner workbook was a new research instrument, so it was essential to establish its validity prior to conducting the full study. This was achieved through the drafting and piloting of the workbook as well as recruiting an expert panel to review the proposed set of calculations. In contrast, since the confidence questionnaire was a well-established research instrument which had been deployed in a previously published peer-reviewed study (Eccles et al., 1993), it did not attract the same validity concerns. Nevertheless, there were potential validity concerns to address regarding the Phase 2 interviews. For example, the ‘interviewer effect’ (Denscombe, 2010, p. 193) could distort the data. The interviewer effect recognised the influence of individual interviewers on their data collection and the interview protocol was intended to moderate those effects. For example, it supported a consistent approach towards each interview. Also, recognising the tendency of some young learners to attempt to please adults with their responses, the interview script which was read aloud to learners stressed that the interviewer was keen to learn about their ideas. The learners’ responses in the pilot study were cross-checked with their teacher who felt that the learners gave honest responses to the set of questions, reinforcing the validity of the questioning route.

Reliability.

If a research design was reliable, then other researchers who applied the same approach should achieve similar findings (Yin, 2009, p. 45). The literature frequently stressed that the reliability of a study depended on the consistency of its data (including Denscombe, 2010, p. 298; Fowler, 2002, p. 77). By stating my approaches towards ensuring the reliability of my data, future researchers should be able to replicate my study.

Two of the potential reliability issues for this study were the timing of the data collection and the research instruments. The potential timing issues were addressed by requiring all of the participating schools to administer both the confidence questionnaires and workbooks the week before the October half term holiday and conducting all of the face-to-face interviews in the week immediately after that break. The reliability of the data collection instruments was addressed in different ways. The confidence questionnaire's reliability was confirmed by testing the consistency of the responses using Cronbach's Alpha test. For further details, please refer to the section titled 'Confidence' in Chapter Four. The clarity of the workbook was confirmed by teachers during the drafting and piloting process. The reliability of the interviews was enhanced by adopting a semi-structured approach with an accompanying interview protocol addressed by sharing the responses from the pilot study with the class.

Research Ethics

Educational research has been defined as "finding the truth" rather than "telling the truth" (Pring, 2000, p. 144). The definition reflected my own understanding of the role of the education researcher as someone who was searching for information rather than simply relaying it. The challenges of the role have been recognised through the development of ethical guidelines by BERA (2011) and my research was conducted in full accordance with its ethical guidelines.

The demand for respecting the confidentiality of participants has been described as the most important ethical concern for researchers (Singer, 2008, p. 96). She drew attention to the need for the researcher to protect their data and the confidentiality of participants. Thus, all data was stored on a password protected laptop and a UIC replaced the real name of each learner. Pseudonyms replaced any real names in the final report.

All of the participants in the study were volunteers. An opt-out procedure, developed on the advice of a head teacher, was adopted because the data was deemed low risk. This approach enabled adults to withdraw their children by signing and returning the consent form (Appendix J). Otherwise, in agreement with their school, their participation would be assumed. This approach considerably reduced the paperwork burden for schools. Furthermore, Robson (2002, p.70) suggested that participating children in a study might be old enough to confirm their agreement too. With that in mind, the interview protocol clarified the purposes of the study for learners at the beginning of their interview as well as their right to withdraw at any time.

At the beginning of this study, I was curious to learn more about gender differences surrounding calculation fluency following the reforms (DfE, 2013a). As the project progressed, I became more aware of my growing unease with the new policy which promoted formal algorithms over and above other strategies. Moreover, I was aware that my potential bias might affect the way in which I analysed and interpreted my data. Although I was a single researcher, I was working within a large mathematics department and liaising with an education faculty too. Therefore, throughout this project, I took every opportunity to present my ongoing work to research groups and conferences in order to receive feedback which might challenge my approaches and interpretations. For example, one conference delegate challenged my decision to categorise learners as boys or girls. Her arguments forced me to reflect on my research design, although I decided to keep both of those categories because they enabled my research to be evaluated against other studies which had adopted the same approach, and I welcomed other feedback throughout the project.

Chapter Four: Phase 1 Results

Introduction

This chapter presents the results from the Phase 1 data collection. The data was analysed with R (R Core Team, 2016). The first section outlines the composition of the Phase 1 sample, which includes its gender balance, prior attainment and confidence levels. This is followed by the statistical analyses of the Phase 1 data in order to address RQ1, RQ2 and RQ3. The chapter concludes with a summary of the main findings from Phase 1 as well as the key questions for further exploration in Phase 2 of this MMR study.

Phase 1 Sample

The Phase 1 sample was drawn from a cohort of 82 schools participating in a regional MaST programme. Each of those 82 schools agreed to participate in the data collection. The cohort covered a wide geographical region across the East and East Midlands. Altogether the region included schools from 19 different Local Authorities (LAs). At that time, each of those LAs organised its own School Improvement Service (SIS) which offered bespoke mathematics training for its schools. Hence my sampling decisions needed to consider both the geographic spread of my MaST population as well as the differing mathematical experiences of schools working with different SISs across the LAs.

For quantitative data collection "Random sampling is perhaps the most well-known of all sampling strategies" (Teddlie & Yu, 2009, p. 79). It required selecting "a subset of cases from the known complete population," (Gorard, 2015, p. 75). In other words, random sampling required choosing a specific number of schools from my wider population of schools. Moreover, random sampling offered very straightforward ways of identifying the sample schools by drawing the school names out of a box or using computer-generated random number tables. Another advantage of random sampling for my study was its suitability for providing, "a less-biased estimate of a more general population than any other kind of sample," (Gorard, 2015, p. 93). However, cluster sampling offered an alternative approach for this study. Cluster sampling would have involved organising a population of into clusters and choosing a random number of those clusters as the sample (Teddlie & Yu, p.79). Hence, one of advantages of cluster sampling for my study was that it might have reduced my travelling since groups of the schools would have been located together. However, cluster sampling had its drawbacks too. Of concern for my own study, a cluster of schools in one area might have experienced very different mathematical support from its LA than schools in a neighbouring area and such variances might not have been apparent in the resulting data set, making it more difficult to generalise my results to the population. Therefore, I discarded cluster sampling in favour of random sampling for Phase 1.

The actual Phase 1 sample consisted of 590 Y6 learners, 324 boys (54.9%) and 266 girls (45.1%) drawn from a potential sample of 670 Y6 learners, 372 boys (55.5%) and 298 girls (44.5%). A total of 80 learners (48 boys, 32 girls) were withdrawn from the study for the following reasons: their prior attainment KS1 scores were unavailable (22 boys, 13 girls); their current attainment was below the level of the workbook calculations (20 boys, 13 girls); they were absent on the day of the task (3 boys, 3 girls); the teacher had behavioural concerns about the learner (1 boy); parental permission was not granted (2 boys, 2 girls); and, learner was newly arrived to UK (1 girl). It should also be noted that one of the participating schools was forced to withdraw from Phase 1 of the study at short notice due to an impending government inspection, but it was replaced with another school of similar size.

Gender.

Although there was a greater proportion of boys than girls in Phase 1, further examination of the national data for the same cohort also revealed a gender imbalance favouring boys. Out of 590,930 eligible Y6 learners nationally, there were 301,682 boys (51.1%) and 289,248 girls (48.9%) (DfE, 2013d). A chi-square test comparing the gender balance in the national data set with own my sample revealed that the gender difference in my sample was not statistically significant, χ^2 (1) = 3.59, p = 0.06.

Prior attainment.

The prior attainment data was sorted into three categories: Below Level 2 (L2) - for learners working at or below Level 1 (L1); L2; or Level 3 (3) (Table 10). Beginning with the data for learners working below L2, there was a larger proportion of the boys than girls but a chi-square test of independence revealed that there was no significant gender difference, χ^2 (1) = 0.08, p = 0.78. The majority of the learners were L2 and there was a higher proportion of L2 girls than boys but a chi-square test of independence indicated that the difference was not significant χ^2 (1) = 0.49, p = 0.49. Approximately a fifth of the learners were working at the higher L3, with a larger proportion of the boys but the difference was not significant, χ^2 (1) = 1.13, p = 0.29.

Table 10

KS1 Scores of the Participating Learners (N = 590)

KS1 Level	Boys (%) *n* = 324	Girls (%) *n* = 266	Total (%) *N* = 590
Below 2	31 (9.6)	23 (8.6)	54 (9.2)
2	216 (66.7)	191 (71.8)	407 (69.0)
3	77 (23.8)	52 (19.5)	129 (21.9)
Total	324 (100)	266 (100)	590 (100)

Confidence.

Each learner completed a confidence questionnaire consisting of 10 questions requiring responses using a Likert-style response scale ranging from one (minimum) to five (maximum). The analysis stage offered an opportunity to assess its reliability by considering its internal consistency. There were several different ways to measure its internal consistency during the analysis, which involved checking the consistency of the results across each of the ten questions and their five possible responses; a Cronbach Alpha test was conducted because it was deemed suitable for questionnaires which included questions addressing different topics (Cohen et al., 2007, p. 148). The confidence questionnaire was deemed highly reliable (10 items; $\alpha = 0.88$) since the reliability of a questionnaire according to Gibson (2010, p. 74) increased as its Cronbach Alpha score approached one.

Exploring the association between confidence levels and gender highlighted that the mean confidence level was higher for the boys ($M = 39.60$, $SD = 6.87$) than the girls ($M = 37.72$, $SD = 7.31$), and an independent samples t-test indicated that there was a significant gender difference in the confidence scores, t (588) = 3.22, $p < 0.01$. In other words, the boys were more likely to report high confidence scores than the girls.

Phase 1 Findings for RQ1

RQ1: To what extent do gender, prior attainment and confidence predict use of the formal algorithm?

An individual learner's formal algorithm score was given by the total the number of times that the learner chose the formal algorithm across the 16 calculations. Hence the formal algorithm scores ranged from a minimum score of zero, for those learners who did not choose any formal algorithms at all, to a maximum score of 16 for those who chose a formal algorithm for each of their calculations.

The data was subjected to multiple regression analysis (MLR). Simple linear regression analysis enabled the prediction of an outcome variable from a single predictor variable (Field, Miles & Field, 2012, p. 198). However, since my data set contained the three variables of gender, confidence and prior attainment, MLR was the more appropriate model. By taking into account a number of contributory factors, the process increased the accuracy of predictions (Gorard, 2001, p. 172). Moreover, one of the main advantages of using MLR was its suitability for providing "an estimate of the relative importance of the different independent variables in producing changes in the dependent variable" (Robson, 2002, p. 431). In other words, MLR analysis would reveal the relative importance of gender, confidence and prior attainment on the use of the formal algorithm. MLR was regarded as "an extremely powerful technique for untangling the complex and interrelated relationships between variables, as is typical in education" (Hartas, 2010, p. 391). However, it also had its limitations. One of the key issues using MLR was the possibility of interaction between the predictor variables which may have been intercorrelated, known as multicollinearity (Cohen, 2007, p. 541). Hence my MLR analysis needed to check whether the level of correlation between my three predictor variables was within acceptable levels. Nevertheless, if multicollinearity was an issue, then it could be addressed by either removing one of the variables from the model or combining them to create a new variable (Gorard, 2001, p. 172).

One of the nine assumptions of MLR (Field et al., 2012, p. 271) was that it required a continuous dependent variable and more than one predictor variable, and the formal algorithm score satisfied the definition of a continuous dependent variable due to the equal intervals along its scale representing equal differences in the variable (Field et al., 2012, p.9; Gorard, 2001, p. 60), However MLR also restricted any categorical predictor variables to two categories (Field et al., 2012, p. 271). Although the gender variable satisfied that requirement, the prior attainment variable had three categories. The issue was resolved by using two dummy variables, the first dummy variable was L2 and the second dummy variable was L3 so a learner working below L2 was defined as 'not being L2' and 'not being L3.' The basic equation for MLR, and simple regression, was given by:

$$\text{outcome}_i = \text{model} + \text{error}_i$$

Since MLR also required a coefficient for each of the additional predictor variables, the following initial regression equation was applied to the data:

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i$$

Note. y_i is the formal algorithm score for the learner i; α is the intercept; x_{i1} is whether or not the learner is a boy; x_{i2} is whether or not a learner achieved Level 2*;* x_{i3} is whether or not a learner achieved Level 3; x_{i4} is the confidence score; β_1 is the regression coefficient for gender; β_2 is the regression coefficient for Level 2; β_3 is the regression coefficient for Level 3; β_4 is the regression coefficient for confidence; ε_i is the error term for learner i.

Using the enter method, otherwise known as forced entry, the predictor variables were simultaneously entered into the model. The entry method was chosen, following the advice of Field et al. (2012, p. 264), because the literature review had not indicated any possible hierarchy for the data entry. Unless the literature had supported a different method, the entry method was the standard method of variable entry. An *F*-test assessed whether the set of predictor variables collectively predicted the use of the formal algorithm and *R*-squared, the multiple correlation coefficient of determination, determined how much variance in the dependent variable could be accounted for by those predictor variables and the *t*-test determined the significance of each predictor and beta coefficients were used to determine the magnitude of prediction for each independent variable (Table 11). The findings resulted in the following statistically significant regression equation (F (4, 585) = 11.98, p<0.01), R^2 = 7.57%:

Use of formal algorithm = 5.52 + 0.89 Gender -1.30 L2 -1.67 L3 + 0.14 Confidence

Table 11

Multiple Regression Coefficients for Predicting Use of Formal Algorithms

Predictor	β	*SE*	*t*
Intercept	5.5212	0.9884	5.5862**
Gender	0.8893	0.3091	2.877**
L2	-1.3026	0.5379	-2.4217*
L3	-1.6688	0.6225	-2.6806**
Confidence	0.1439	0.0229	6.2793**

Note. Gender code for girl = 1, boy = 0; L2 =1, other levels = 0; L3 =1, all other levels = 0; confidence ranges between 10 to 50.
* p < .05. ** p < .01.

Table 11 revealed that all of the three predictor variables were significant predictors of the formal algorithm score, but confidence was the most significant predictor variable. More specifically, increased confidence was related to increased use of the formal algorithm. An individual's use of the formal algorithm increased by 1.44 for every ten-unit increase in their mathematical confidence. In other words, learners with the maximum score of 50 were predicted to choose the formal algorithm for six calculations more than learners with the minimum confidence score of ten. In contrast, there was an inverse association between the levels of prior attainment and use of the formal algorithm. L2 learners were likely to use the formal algorithm for approximately one less calculation than those learners working below L2. For L3 learners, the difference was almost two calculations less than learners working below L2. It also revealed the significant role of gender when choosing the formal algorithm, even when controlling for confidence and prior attainment. Moreover, it predicted that girls tended to choose the formal algorithm for one more calculation than the boys.

However, two key points should be noted about the regression model. First, since confidence had a higher significance for predicting use of the formal algorithm than the other two predictor variables, its role in the regression model was examined in more depth. Hence I ran a model exploring whether confidence had an additional non-linear (namely, quadratic) relationship to each dependent variable, but the result was non-significant. Hence the initial model was accepted for predicting use of the formal algorithm. The second key point addressed the low R^2 value which indicated that just under 8% of the variation in the use of the formal algorithm score was explained by taking into account gender, prior attainment and confidence. Phase 2 offered an opportunity to explore the other factors responsible for the remaining 92% of the variance.

Checking assumptions.

Berry (as cited in Field et al., 2012, pp. 271-272) noted that MLR required that several assumptions should be satisfied, including the assumptions requiring quantitative or categorical predictor variables with a quantitative, continuous and unbounded dependent variable, which were satisfied earlier in this section. The non-zero variance requirement was satisfied since none of the three predictor variables had variances of zero and the multicollinearity requirement was satisfied by checking that the level of correlation between each pair of independent variables was within acceptable limits. As mentioned previously, the multicollinearity requirement was one of the key issues when using MLR. Hence, I needed to check that the level of correlation between each pair of my three independent variables was within acceptable limits. Although the three variables satisfied this requirement by falling within the acceptable limits, the assumptions of MLR made allowances for a certain level of multicollinearity; Field et al. (2012, p. 214) recommended setting the maximum acceptable value as 0.8, noting that a value of 1 would have indicated a perfect score and a linear association between them and, since none of the pairs for confidence, gender or prior attainment exceeded 0.8, this assumption was satisfied for my predictor variables (Appendix M). The remaining four assumptions related to linearity, homoscedasticity, normality, and influential outliers. The linearity assumption had two distinct components: the linear relationship, which was required between the dependent and independent variables, was satisfied by examining the scatterplots for an approximately straight-line relationship between the predictor variables and the dependent variable; the requirement for linear relationship between the dependent variable and the non-categorical independent variables was satisfied using a visual inspection of the partial regression plot. The homoscedasticity assumption, which required that the residuals were equal for all values of the predicted dependent variable, was satisfied by visually confirming an approximately horizontal spread of the residuals along the range of the predictor variables. The requirement for normally distributed errors was satisfied by examining the Normal Q-Q plot which confirmed that the residuals closely followed the dotted normality line. Furthermore, the potential influence of the outlier values was within acceptable limits since all cases were well within Cook's Distance whose dotted line fell outside the plotted area.

Summary.

The MLR model for RQ1 indicated that confidence, gender and prior attainment were all significant predictors of the use of the formal algorithm amongst Y6 learners. The most significant factor was confidence; learners with high confidence scores were predicted to have higher use of formal algorithms than low confidence learners. Gender was also a significant factor for predicting use of the formal algorithm, the girls were significantly more likely than the boys to choose formal algorithms. The significant association between prior attainment and use of formal algorithms was an inverse association, learners with higher prior attainment scores were less likely to use formal algorithms than learners with lower prior attainment scores. In Phase 2, the underlying reasons behind these findings were explored in more depth.

Phase 1 Findings for RQ2

RQ2 To what extent do gender, prior attainment or confidence predict calculation accuracy?

The calculation accuracy of the learners was measured by their total number of correct answers for the 16 calculations, so their possible scores ranged from zero to 16. To examine the extent that gender, prior attainment and confidence predicted their calculation accuracy, MLR was applied to the data. As before, the three predictor variables were gender, prior attainment and confidence but the dependent variable for RQ2 was accuracy. Since calculation accuracy was calculated in equal intervals representing equal differences in the variable, the accuracy score satisfied the MLR requirement for a continuous dependent variable (Field et al., 2012, p.9 and Gorard, 2001, p. 60). The following regression equation was used:

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 \mathrm{x}_{\mathrm{i}4} + \varepsilon_i$$

Note. y_i is accuracy for learner *i*; α is the intercept; x_{i1} is whether or not the learner is a boy; x_{i2} is whether or not a learner achieved Level 2; x_{i3} is whether or not a learner achieved Level 3; x_{i4} is confidence; β_1 is the regression coefficient for gender; β_2 is the regression coefficient for Level 2; β_3 is the regression coefficient for Level 3; β_4 is the regression coefficient for confidence; ε_i is the error term for learner *i*.

As with RQ1, the predictor variables were simultaneously entered into the model using the entry method. A statistically significant regression equation was found (F (4, 585) = 74.36, p<0.01), R^2 = 33.71% (Table 12):

Accuracy = -0.528 + 0.030 Gender + 2.926 L2 + 5.328 L3 + 0.187 Confidence

The most significant factors affecting accuracy were prior attainment and confidence, but not gender. There was a significant direct association between confidence and accuracy. For every unit increase in confidence, there was a corresponding 0.19 increase in accuracy. If a learner reported a maximum confidence of 50, then the model predicted that they would correctly answer around seven more calculations than a learner with the minimum confidence score of ten. More confident learners tended to answer more calculations correctly than less confident learners. There was also a significant direct association between prior attainment and accuracy. Learners with a prior attainment of L2 were likely to get three more calculations correct than those learners with a prior attainment below L2. Furthermore, learners with the higher prior attainment of L3 were predicted to correctly answer five more calculations than those learners with a prior attainment below L2. However, in contrast to the findings for RQ1 relating to use of the formal algorithm, gender was not a significant predictor of calculation accuracy. Nevertheless, the MLR model for calculation accuracy only accounted for around 32% of the variance in accuracy. As before, I ran an additional model exploring whether confidence had an additional non-linear (namely, quadratic) relationship to each dependent variable, but it was non-significant. Phase 2 of this MMR study enabled the exploration of other possible factors relating to calculation accuracy.

Table 12

Multiple Regression Coefficients for Predicting Calculation Accuracy

Predictor	β	*SE*	*t*
Intercept	-0.5278	0.8336	-0.6332
Gender	0.0301	0.2616	0.1151
L2	2.9256	0.4516	6.4787**
L3	5.3279	0.5237	10.1749**
Confidence	0.1874	0.0194	9.6621**

Note. G = girl, gender code for boy = 0, for girl = 1; L2 = Level 2, prior attainment code for L2=1, other levels = 0; L3 = Level 3, prior attainment code for L3=1, all other levels = 0; confidence ranges between 10 to 50.
* p < .05. ** p < .01.

Checking assumptions.

As with RQ1, MLR required the data to satisfy several key assumptions (Field et al., 2012, pp. 271-272). The assumption requiring quantitative or categorical predictor variables with a quantitative, continuous and unbounded dependent variable was satisfied earlier in this section. The non-zero variance requirement was satisfied since none of the three predictor variables had variances of zero. As before, the multicollinearity requirement was satisfied by confirming that the level of correlation between each pair of independent variables was below the recommended maximum score of 0.8 (Field et al., 2012, p. 214). The remaining four assumptions related to linearity, homoscedasticity, normality, and influential outliers. A linear relationship between the dependent and independent variables was satisfied by examining the scatterplots for an approximately straight-line relationship between the predictor variables and the dependent variable, and, the requirement for linear relationship between the dependent variable and the non-categorical independent variables was satisfied using a visual inspection of the partial regression plot. The homoscedasticity assumption was satisfied by visually confirming an approximately horizontal spread of the residuals along the range of the predictor variables. The requirement for normally distributed errors was satisfied by examining the Normal Q-Q plot which confirmed that the residuals closely followed the dotted normality line. Furthermore, the potential influence of the outlier values was within acceptable limits since all cases were well within Cook's Distance whose dotted line fell outside the plotted area.

Summary.

The MLR analysis results indicated that a learner's confidence and their prior attainment, but not gender, were significant predictors of their calculation accuracy in Y6. Learners with high confidence scores were also likely to have high accuracy scores. However, the quantitative data did not explain how a more confident learner's approach to their calculations resulted in more accurate answers than a less confident learner. Likewise, although the findings indicated that learners with high prior attainment scores tended to achieve high accuracy scores, they did not explain how those learners achieved those higher accuracy scores. Hence, due to t he MMR nature of this study, those findings were explored in more depth during Phase 2.

Phase 1 Findings for RQ3

RQ3: Which are the most accurate calculation strategies for Y6 learners completing age-related, context-free multi-digit written calculations?

The results showed that the most frequently selected strategies were not necessarily the most accurate strategies. The accuracy rates for the strategies chosen for each of the 16 calculations are presented in Appendix K.

Most frequently selected strategies.

The formal algorithm was the most frequently selected strategy for each of the 16 calculations. Moreover, the proportion of learners who chose the formal algorithm for addition and subtraction calculations was higher than the proportion who chose it for multiplication and division calculations. These findings verified similar findings by Borthwick and Harcourt-Heath (2007). Table 13 presents the strategies chosen by at least ten of the learners for each of the 16 calculations.

Table 13

Strategies Chosen by at least 10 of the Learners for each of the 16 Calculations (N = 590)

	FA	P	NA	R	CB	CD	CU	M10	AO	G	E	NB	O
456+372	544	34	-	-	-	-	-	-	-	-	-	-	12
5412+2584	546	24	-	-	-	-	-	-	-	-	-		20
245+256	540	29	-	-	-	-	-	-	-	-	-	-	21
299+532	520	22	22	13	-	-	-	-	-	-	-	-	13
632-154	502	16	-	-	16	-	35	-	-	-	-	-	-
500-76	401	-	-	-	63	-	53	-	41	-	-	13	-
702-695	444	13	31	32	20	-	28	-	19	-	-	-	-
382-199	473	14	21	22	10	-	38	-	10	-	-	-	-
20x46	255	48	-	-	-	-	-	31	-	180	48	-	28
27x63	258	53	22	-	-	-	-	-	-	185	55	-	23
35x99	248	46	-	28	-	-	-	-	-	199	46	-	-
568x34	254	41	52	-	-	-	-	-	-	180	45	-	18
693÷3	323	41	101	-	-	29	19	-	46	-	-	-	31
480÷20	255	18	84	-	-	60	65	19	50	-	-	-	39
401÷25	252	-	94	0	0	145	20	0	50	-	0	0	34
517÷19	235	19	145	0	0	81	32	0	33	-	0	0	40

Note. FA = formal algorithm; P = partition; NA = no answer; R = round; CB = count back; CD = chunk down; CU = count-up; M10 = multiples of ten; AO = answer only; G = grid method; E = expanded; NB = number bonds; O = other.

Most accurate strategies.

The accuracy rates for each of the strategies attempted for each individual calculation were worked out by dividing the number of correct responses for each strategy by the total number of attempts using that strategy. Although the formal algorithm was the most frequently selected strategy for each calculation, the data also revealed that it was only the most accurate strategy for the calculations 456 + 372, 5412 + 2584, 299 + 532 and 693 ÷ 3 (Table 14). Two of those calculations (456 + 372 and 5412 + 2584) were included in the research design due to their deemed suitability for using formal algorithms, but the other two calculations were chosen for their suitability for using other strategies (299 + 532 and 693 ÷ 3); yet the learners were more accurate using the formal algorithm for both calculations.

Table 14 illustrated that other strategies were more accurate than formal algorithms for the three-quarters (12) of the 16 calculations, including three calculations which were included in the research design due to their perceived suitability for using a formal algorithm (632 - 154, 27 x 63 and 517 ÷ 19). Nine strategies other than formal algorithms featured in the list of most accurate strategies and three of those featured twice; rounding (382 – 199 and 35 x 99), counting-up (632 – 154 and 401 ÷25) and chunking down (480 ÷ 20 and 517 ÷ 19)

Table 14

Strategies with the Highest Number of Correct Answers for Each Calculation

	Calculation Strategy									
Calculation	Formal algorithm	Partition	Count back	Number bonds	Count-up	Round	Multiples of 10	Grid	Expanded	Chunk down
456 + 372	✓									
5412 + 2584	✓									
245 + 256		✓								
299 + 532	✓									
632 - 154			✓							
500 - 76				✓						
702 - 695					✓					
382 - 199						✓				
20 x 46							✓			
27 x 63								✓		
35 x 99						✓				
568 x 34									✓	
693 ÷ 3	✓									
480 ÷ 20										✓
401 ÷ 25					✓					
517 ÷ 19										✓
Total	4	1	1	1	2	2	1	1	1	2

The next step in the analysis was assessing whether the most accurate strategies for each calculation were significantly more accurate than the other strategies attempted by the learners. A series of one-way ANOVA tests was conducted to compare the accuracy scores for the strategies chosen for each calculation to assess whether their differences were significant. The advantage of using ANOVA tests was their suitability for comparing the accuracy scores of more than two strategies for each calculation. The data satisfied the assumptions of ANOVA testing (Field et al., 2012, p. 359): the mean accuracy scores of at least three independent calculation strategies were considered for each of the 16 calculations; the data satisfied the requirements for a continuous, normally distributed dependent variable and independent categorical variables which had no significant outliers; and, for calculations where less than five learners had chosen a particular strategy, the results were combined into the ‘other’ category in order to satisfy the minimum data requirements for ANOVA. However, examining the data at the level of individual calculations had potential consequences regarding errors. In the methodology section, attention was drawn to the issue of clustering, noting that my findings should be regarded as indicative rather than conclusive. It was inadvisable to conduct multiple tests on subsets of a dataset without acknowledging the possibility of achieving a false positive, known as a Type 1 error in the results (Gorard, 2001, p. 151). To avoid such errors during the analyses, the *p*-values were corrected using the Bonferroni correction $\alpha/N = 0.003$ (Howitt & Cramer, 2005, p. 249). Although an ANOVA test could reveal whether there were any significant differences in the means, one of its disadvantages was that ANOVA could not identify which means were significantly different since it was an omnibus test (Field et al., 2012, p. 400). Therefore, any significant ANOVA results were followed up with Tukey *post hoc* tests which identified the actual strategies with statistically significant means (Gorard, 2001, p. 160).

Whilst the earlier findings had shown that formal algorithms were the most frequently selected strategy for every calculation, and that they were only the most accurate strategy for four of the 16 calculations, ANOVA testing and the subsequent Tukey *post hoc* tests (Appendix I) revealed that formal algorithms were only significantly more accurate than other strategies for two of the 16 calculations; 456 + 372, $F(3, 586) = 13.36$, $p<0.001$ and 693 ÷ 3, $F(7, 582) = 79.39$, $p<0.001$. Further, Cohen’s (1988) convention for effect sizes suggested that their effect sizes were medium ($d = 0.56$) and medium-to-large ($d = 0.67$) respectively. However, other strategies were significantly more accurate than formal algorithms for four of the 16 calculations. Counting-up was significantly more accurate than the formal algorithm for calculating 702 - 695, $F(6, 583) = 18.67$, $p<0.001$ and the effect size was large ($d = 0.85$). The grid method was significantly more accurate than long multiplication for calculating 27 x 63, $F(6, 583) = 14.38$, $p<0.001$ but the effect size was small-to-medium (0.34). Chunking down was significantly more accurate than long division for calculating both 401 ÷ 25, $F(7, 582) = 23.38$, $p<0.001$ and 517 ÷ 19, $F(6, 583) = 28.28$, $p<0.001$, and the effect sizes were medium ($d = 0.45$) and medium-to-large ($d = 0.70$) respectively.

It should also be noted that accuracy varied by operation. The learners achieved much higher scores for addition (90.6% correct) than subtraction (52.2 % correct), multiplication (44.4 % correct) or division (44.4 % correct). These findings were also followed up in the qualitative analysis.

Summary.

Although the workbooks deliberately contained some calculations which were deemed more appropriate for using formal algorithms than others, the formal algorithm was the most frequently selected calculation strategy for each of the 16 calculations. However, it was only significantly more accurate than the other strategies chosen by the learners for just two of those 16 calculations whereas other strategies were significantly more accurate than formal algorithms for four of the other calculations. Due to the explanatory nature of this study, the reasons behind learners choosing the formal algorithm for each calculation were addressed in the subsequent qualitative Phase 2 as well as the differing accuracy rates for the four operations.

Phase 1 Summary

The Phase 1 analysis highlighted the crucial role of a learner's confidence when performing written calculations. Confidence was the most significant variable affecting both accuracy and use of the formal algorithm. More confident learners tended to work more accurately. Also, more confident learners were more likely to choose the formal algorithm than less confident learners. However even after constructing models to predict an individual learner's use of the formal algorithm or accuracy rates, most of the variance in those models remained unexplained. The sequential explanatory nature of this MMR study allowed the reasons behind these findings to be examined in the following qualitative phase.

Prior attainment also played a significant role in calculation fluency, but it was not as significant as confidence; higher attaining L3 learners tended to work more accurately than the other learners. Also, there was an inverse association between prior attainment and use of the formal algorithm; higher attaining L3 learners were less likely to use the formal algorithm than L2 learners.

Gender played a less significant role in calculation accuracy, or use of the formal algorithm, than either confidence or prior attainment. Nevertheless, the finding that girls were significantly more likely than the boys to choose the formal algorithm was also followed up in Phase 2. There were no significant gender differences relating to calculation accuracy.

The formal algorithm was the most frequently selected strategy for each of the 16 calculations even though several of the calculations had been included due to their deemed suitability for using other strategies. Furthermore, the formal algorithm was only the most accurate strategy for just two of the 16 calculations. The reasons behind the strategy choices of the learners were also explored in the qualitative Phase 2.

Chapter Five: Phase 2 Findings

The three key aspects of Phase 2 which reflected the sequential explanatory nature of this mixed methods study were the results arising from the Phase 1 analysis directly informing the sampling for Phase 2 data collection, the Phase 2 analysis offering the opportunity to identify further variables to explain the Phase 1 findings and Phase 2's data collection and analysis addressing the two remaining aspects of the calculation fluency from Phase 1.

Phase 2 Sample

Due to the explanatory nature of this study**,** a purposive subsample of Y6 learners (Table 15) was drawn from the larger Phase 1 sample (N = 590). The procedures for identifying the interviewees were outlined in Chapter Three, including changing both their names and their school names in accordance with the ethical approaches adopted for this study.

Table 15

Phase 2 Interviewees and their Phase 1 data (n = 23)

Name	School	Gender	Level 3	Low confidence	Total score	Total use of formal algorithm
Alex	A	Boy	Yes	No	16	10
Ben	A	Boy	Yes	No	11	11
Claire	A	Girl	Yes	No	12	12
Daisy	A	Girl	Yes	No	13	9
Gary	A	Boy	No	Yes	13	12
Harry	A	Boy	No	Yes	10	14
Milly	A	Girl	No	Yes	11	12
Lucy	A	Girl	No	Yes	11	11
Edward	B	Boy	Yes	No	15	10
Fred	B	Boy	No	Yes	13	7
Amy	B	Girl	Yes	No	11	9
Beth	B	Girl	No	Yes	9	5
James	B	Boy	Yes	No	12	11
Kevin	B	Boy	No	Yes	13	7
Kate	B	Girl	Yes	No	4	8
Jessica	B	Girl	No	Yes	11	16
Charlie	C	Boy	Yes	No	10	8
David	C	Boy	No	Yes	12	5
Emily	C	Girl	Yes	No	15	10
Frans[a]	C	Girl	No	Yes	12	10
Liam	C	Boy	Yes	No	9	4
Mark	C	Boy	No	Yes	9	7
Gabby	C	Girl	No	Yes	10	12
Helen	C	Girl	Yes	No	9	13

Note. Low confidence = confidence score was in the lowest quartile; Total Score = individual's total score out of 16; Total use of formal algorithm = number of calculations using the formal algorithm.

[a] Learner was unexpectedly absent due to illness on interview day.

Preparing the Framework Analysis Grids

Chapter Three described the steps taken during the qualitative analysis, including the development of the themes and sub-themes for the framework analysis. The themes related to the various aspects of calculation fluency, featuring sub-themes as well as their descriptions for the subsequent coding process. Figure 12 presents the sub-themes for the analysis of the qualitative data relating to use of the formal algorithm.

Theme: Use of the formal algorithm	
Sub-Theme	Description
Use of multiple strategies	Awareness of more than one strategy
	Willingness to use other strategies
Role of SATs	Justifies strategy selection based on SATs

Figure 12. Sub-themes for the 'Use of the formal algorithm' analysis theme.

Using the themes and their sub-themes, the interview transcripts were coded - or indexed - in readiness for the charting process. During coding, the transcript for each learner was annotated to highlight the various themes and sub-themes within it. Figure 13 presents an excerpt from Charlie's coded transcript for the 'Use of the formal algorithm' theme.

Line	Speaker	Transcription
275	Interviewer	This is 480 divided by 20 and you've done that one by…
276	Charlie	The bus stop
277	Interviewer	Any other way you could have done that?
278	Charlie	Could have done it like by counting on my fingers or in my head
279	Interviewer	Counting up in 20s?
280	Charlie	Yeah, or halve 480.
281	Interviewer	If you halve it, what do you need to do to get the answer?
282	Charlie	Erm…just, erm…take away a zero.
283	Interviewer	So, you've got more than one way to do it…
284	Interviewer	Which is the best way of doing it would you say?
285	Charlie	In my head but when it says 'Show your method' I usually do it
286	Charlie	[formal algorithm] so I get the marks for my working out.

Figure 13. Excerpt from Charlie's coded transcript for the 'Use of the formal algorithm' theme.
Note. Green indicates sub-theme 'use of multiple strategies,' blue indicates sub-theme 'role of SATs.'

This process was followed by charting the data which required shifting the focus from working with individual transcripts towards analysing the whole data set. Figure 14 presents an excerpt from the chart for the theme 'Use of the formal algorithm.'

Learner	Calculation	Strategy	Use of multiple strategies	Role of SATs
Charlie	456 + 372	A	Can do it mentally and knows number line but prefers speed of formal I just find it quicker [line 7]	Where it says 'Show your method ' I usually do it [formal] so I get the marks for working out [lines 285-286]
	5412 + 2584	A	As above	
	245 + 256	A	As above	
	299 + 532	A	As above I just do it [formal] to make sure because I know when I do it like I get that right [lines 86-87]	
	500 -76	C	Knows formal but reluctant to use it because lacks practise: if I remember it [line 115]	
	632 – 154	C		
	382 – 199	C		
	702 - 695	C		
	20 x 46	G	Knows long multiplication but lacks confidence with it when I was doing that in Year 5 I didn't get a bit of it [line 192]	
	35 x 99	G	I could have just round the 99 to a 100 and taken 35 away [line 234]	
	27 x 63	G		
	568 x 34	G		
	401 ÷ 25	A	Does not know other written ways It's just what I know [line 253]	
	480 ÷ 20	A	Counting on my fingers or in my head [line 278] Yeah or halve 480 [line 280]	
	693 ÷ 3	A	Did not consider the numbers before applying formal I just normally write it down, I didn't really notice [the numbers] [line 293]	
	517 ÷ 19	A		
David	456 + 372	E	It's easier because you do 400 and 300 (700), 50 and 70 (120) and then 2 and 6 (8) and then add them together [lines 8 and 9] Also knows formal algorithm but did not consider using it except partitioning is quicker I suppose [line 22]	
	5412 + 2584	E		
	245 + 256	E	That one you might do in your head [line 39]…yer, easily [line 65]	No, it says do your working out [line 68]
	299 + 532	E		

Figure 14. An extract from the 'Use of the formal algorithm' chart.

Note. Underlining indicates verbatim text.

The following sections present the qualitative findings arising from the framework analysis, organised by RQ.

Phase 2 Findings for RQ1

RQ1: To what extent do gender, prior attainment and confidence predict use of the formal algorithm?

Although the Phase 1 findings for RQ1 revealed that prior attainment, confidence and gender were statistically significant variables for predicting calculation flexibility, the three variables were insufficient to explain the variance in the data. Phase 2 offered an opportunity to deepen our understanding about the calculation choices of the interviewees.

Descriptive statistics for the interviewees relating to RQ1.

The descriptive statistics for the interviewees mirrored the results for the wider Phase 1 sample in several ways. More specifically, the low confidence interviewees chose formal algorithms more frequently (*M* = 9.8, *SD* = 3.8) than the higher attaining learners (*M* = 9.6, *SD* = 2.3). Also, the girls chose formal algorithms more often (*M* = 10.6, *SD* = 2.9) than the boys (*M* = 8.8, *SD* = 3.0).

However, the interviewees did differ from the wider sample when considering their most frequently selected multiplication strategies. Across the Phase 1 sample, formal algorithms were the most frequently selected strategies for each of the 16 calculations and that was also the case for the interviewees for three out of the four operations, but the grid method was their most frequently selected multiplication strategy. The use of the formal algorithm by each of the 23 interviewees in their workbooks is presented in Table 16.

The interviewees were asked to explain why they had chosen their strategies and whether they knew any alternative calculation strategies for their calculations. Their responses are explored in more depth in the following sections.

Use of multiple strategies.

The framework analysis revealed that three-quarters of the learners (9 boys, 9 girls) admitted either not knowing multiple strategies or, if they were aware of them, they either did not recognise the opportunity to apply them or they were reluctant to do so. These findings will be explored in more depth, beginning with the low confidence learners.

Table 16

Use of the Formal Algorithm by the Interviewees (n=23)

	Calculation															
Learner	456 + 372	5412 + 2584	245 + 256	299 + 532	500 - 76	632 - 154	382 - 199	702 - 695	20 x 46	35 x 99	27 x 63	568 x 34	401 ÷25	480 ÷20	693 ÷3	517 ÷19
Alex	✓	✓	✓	✓		✓	✓						✓	✓	✓	✓
Ben	✓	✓	✓	✓		✓	✓	✓					✓	✓	✓	✓
Charlie	✓	✓	✓	✓									✓	✓	✓	✓
David									✓	✓	✓	✓			✓	
Edward	✓	✓		✓	✓	✓	✓						✓	✓	✓	✓
Fred	✓	✓				✓				✓			✓		✓	✓
Gary	✓	✓	✓	✓	✓	✓	✓	✓					✓	✓	✓	✓
Harry	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓
James	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓				
Kevin	✓	✓	✓	✓		✓	✓	✓								
Liam	✓	✓	✓	✓												
Mark	✓	✓	✓	✓									✓		✓	✓
Amy	✓	✓	✓	✓	✓	✓	✓	✓					✓			
Beth	✓	✓	✓	✓		✓										
Claire	✓	✓	✓	✓	✓	✓	✓	✓					✓	✓	✓	✓
Daisy	✓	✓	✓	✓	✓	✓		✓					✓		✓	✓
Emily	✓	✓	✓	✓	✓	✓	✓	✓							✓	
Gabby	✓	✓	✓	✓	✓	✓	✓	✓					✓	✓	✓	✓
Helen	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			✓
Jessica	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kate	✓	✓	✓	✓	✓	✓	✓	✓								
Lucy	✓	✓	✓	✓		✓	✓	✓					✓	✓	✓	✓
Milly	✓	✓	✓	✓	✓	✓	✓	✓					✓	✓	✓	✓

Note. A tick indicates that the learner used the formal algorithm for the calculation.

Eight of the low confidence learners (5 boys, 3 girls) lacked awareness of multiple strategies for some of their calculations. Although most of the group could name alternative addition and subtraction approaches, the majority of them struggled to name alternative ways to multiply or divide their numbers. When Lucy was asked if she could name another approach for division she replied, “No, not really. In Year Four we learnt the bus stop.” Gary offered a similar response, “Erm… Well, I don’t know any other methods for division. It's the only way I know." James did not even attempt any of the division calculations. Similarly, there was a lack of awareness of alternative multiplication approaches for some of the learners. "It's the only way I know," admitted Liam.

In contrast to the low confidence learners, most of the higher attaining learners were aware of multiple strategies. However, half of them (4 boys, 2 girls) did not use those strategies in the workbook task. There were two main reasons for their actions; they either failed to recognise the opportunity to attempt an alternative strategy, or they were reluctant to do so. Both Ben and Charlie realised during their interviews that they could have calculated 702 – 695 using the counting-up strategy. Neither boy appeared to focus on looking on the numbers in a calculation before committing themselves to its calculation. "I didn't really think about it," admitted Ben. Likewise, Charlie confessed, "I didn't really notice how close the numbers were until I noticed [just now] what numbers they were." Edward was another higher attaining learner who admitted overlooking calculations suitable for using alternative strategies. When calculating 382 - 199 he admitted, "I have just seen that it is 199 and I used column there…" Edward was much more reluctant to use multiple strategies when subtracting than adding. For the calculation 500 – 76 he was determined to apply the formal algorithm, "Whatever I'm doing, subtraction is always this method." Alex was another higher attaining boy who was reluctant to deviate from formal algorithms. Looking at the calculation 401 ÷ 25, he rejected using a counting strategy in favour of the formal algorithm, "I've been taught like that." Likewise, he dismissed using near doubles for the calculation 245 + 256, "Probably not because I've been taught column." In contrast, looking at the calculation 401 ÷ 25 which he attempted using the formal algorithm, a surprised Ben blurted out, "I've just realised how I could have done it!" He went on to explain how he could have counted up in steps of 25, admitting it was a much easier approach than using the formal algorithm. For the calculation 382 - 199, Emily was asked whether she had considered using rounding and admitted, "I didn't think of it." Meanwhile, Amy rejected using near doubles for calculating 245 + 256, "I did look at it at the start and I just thought 'I don't really know about this'…"

Role of SATs.

A fifth of the learners (4 boys, 1 girl) admitted that their calculation decisions were influenced by the SATs. The group included four low confidence learners and one higher attaining learner. The decision-making of low confidence learner Fred was clearly influenced by his knowledge of the mark scheme. For the calculation 693 ÷ 3 he noted, "I knew the answer, but I had to write it down." Fred appeared to assume that the only acceptable strategy was the formal algorithm. Likewise, when David rejected using a rounding strategy for a particular calculation, he pointed out, "It says to do your working out." Milly also justified her choice of the formal algorithm rather than adopting a different approach, noting "It says 'Show your working out.'" Kevin was adamant that he was not allowed to use his knowledge of number bonds to make 500 in order to calculate 245 + 256. Likewise, he rejected a mental approach for working out 702 – 695 by explaining, "You have to write it down to get a good mark." Charlie was the only higher attaining learner to justify his calculation choices based on the SATs, "Where it says, 'Show your method' I normally do it [formal algorithm] so I get marks for working out." Their comments revealed a tendency by some learners to incorrectly assume that they needed to use formal algorithms to achieve full marks. In fact, any correct answers would have been awarded full marks in the SATs, but half marks might have been awarded for incorrect attempts using only formal algorithms.

Integrating the findings for RQ1.

Whilst the quantitative results revealed that gender, confidence and prior attainment were all significant predictors of the use of formal algorithms, a large amount of the variance remained unexplained by the regression model. The qualitative findings revealed the importance of knowing multiple strategies, alongside a willingness to use them, and the role of SATs in choosing whether to use formal algorithms.

The quantitative data also showed that the higher attaining learners were less likely to choose formal algorithms than other learners, but the interviews revealed that the higher attaining learners showed a greater awareness of multiple strategies than the others but may not have recognised the opportunities to apply other strategies or may have been reluctant to choose them. Moreover, the quantitative results showed that the low confidence learners were less likely to choose formal algorithms than their more confident counterparts. Their interviews revealed that three-quarters of those low confidence learners lacked knowledge of multiple strategies, often struggling to name multiple approaches for multiplication or division.

The calculation choices of several of the learners appeared to be affected by their knowledge of SATs. Those learners were unwilling to vary their strategy choices away from formal algorithms in the mistaken belief that they had to use that strategy.

Phase 2 Findings for RQ2

RQ2: To what extent do gender, prior attainment and confidence predict calculation accuracy?

The Phase 1 findings revealed that prior attainment and confidence, but not gender, were statistically significant variables for predicting calculation accuracy. The regression model for RQ1 indicated that the higher attaining Y6 learners tended to achieve higher accuracy scores than other learners, whereas learners with low confidence levels tended to achieve lower accuracy scores than the others. Nevertheless, most of the variance in the accuracy scores remained unexplained. The interviews offered an opportunity to explore the reasons behind those findings with the learners themselves. Clearly there were further variables relating to calculation accuracy which were explored during the qualitative analysis.

The accuracy scores for the 23 interviewees are presented in Table 15. The mean accuracy score was higher for the 11 higher attaining learners (M = 11.4, SD = 3.3) than the 12 low confidence learners (M = 11.3, SD = 1.5), mirroring the results of the initial quantitative sample (N = 590). The interview responses highlighted the importance of practising their first-choice strategy, the perceived importance of calculation speed, the role of adults in their decision-making and their checking procedures. Each of those four variables will be addressed in more depth, beginning with practising their strategies.

Practising their strategies.

The interviews revealed the importance of practising their strategies for calculation accuracy. All of the 23 interviewees chose at least one of their strategies because they had practised it. However only two of them, both higher attaining learners, chose all of their strategies because they had practised them. Those two learners, Alex and Daisy also achieved above average accuracy scores. During his interview, Alex discussed practising the grid method approach for multiplication, "[It's] the same with the column method, I've been doing it a long time." Similarly, Daisy noted that she had been introduced to the grid method in Y2. She added that she felt that the formal algorithm for division was easy, "because we've been learning it for quite a while so I've had quite a bit of practice."

Varying accuracy rates for addition, subtraction, multiplication and division.

The accuracy results for the quantitative sample (N = 590) revealed that accuracy varied by operation, the learners achieved much higher scores for addition (90.6% correct) than subtraction (52.2 % correct), multiplication (44.4 % correct) or division (44.4 % correct). Due to the MMR research design, those findings were examined in more depth during the interviews. All of the learners stated that they practised their preferred addition strategies and just one learner (a girl) highlighted her lack of practice with her chosen subtraction strategy. In contrast, just under half of the interviewees (3 boys, 8 girls) said that they practised either their multiplication or division strategies. In fact, Kate struggled with all of her calculations except for the addition questions. She confused subtraction with addition, "I don't know that as well [as addition] and because we are doing about add and take aways sometimes I get muddled up" (see Figure 15).

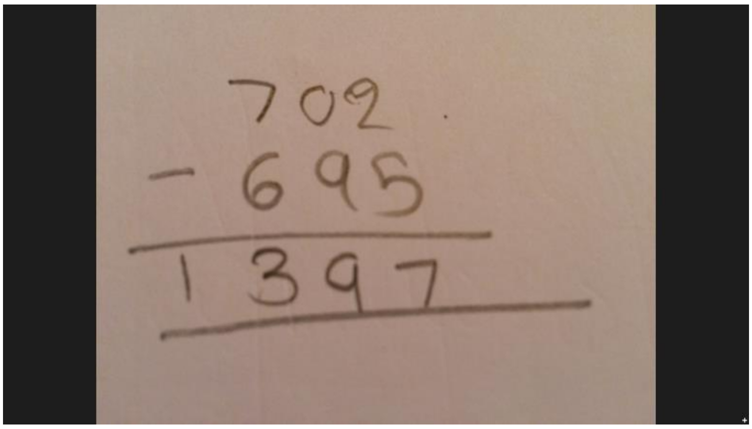


Figure 15. Kate confused subtraction with addition.

None of the low confidence boys practised either their chosen division or multiplication strategies compared with three of the higher attaining boys. One of those higher attaining boys, Edward, reflected on his preference for the formal algorithm for division:

> It's my favourite method to use, like no other methods so far have even come close for me. I enjoy doing it more because it's simple and if you're not sure on an answer you can always do it again. You can look back, see what you did wrong. When you do that, it's not too hard to get hold of and understand.

However, a total of eight higher attaining and low confidence girls practised their chosen multiplication or division strategies. Four of those girls were higher attaining learners, including Amy who chose the formal algorithm for division, "I think we did it in previous years and I forgot. And, we've recently recapped it and it's much easier than I originally thought." Likewise, low confidence learner Jessica found the division formal algorithm easy to use. She explained her approach for calculating 401 ÷ 25, "It's quite simple because you have to try and do 25 into 4 and you can't do it. So, you do 25 into 40 and you can do it once." Nevertheless two-thirds of the low confidence learners (5 boys, 3 girls) appeared to lack confidence applying either their chosen multiplication or division strategies. "We've only been learning division and the bus stop this week," commented Fred. Kevin also struggled with division and did not answer any of those four calculations. Meanwhile Mark seemed unsure which approach he should choose for dividing, "I know the bus stop method so I thought 'Why don't I give that a go?' and it came up a bit wrong…" Harry was the only boy who admitted a lack of confidence with his multiplication strategies, attempting three different strategies across the four calculations. "Yeh, I struggle with that," he admitted. Lucy also struggled with multiplication. Using the grid method for the calculation 35 x 99 she only managed to draw the grid before calculating one of the partial products incorrectly. "I got a bit confused on that one," she reflected. Milly and Helen both noted their lack of practice with the division formal algorithm. Helen admitted “I’m not very good at that one” whereas Milly noted that, “When we did this I didn’t know how to do long division.”

Checking procedures.

During the analysis of students’ workbooks, it was apparent that some of the learners might have checked their answers. For example, some of the scripts indicated that a learner had attempted a second strategy which might have been an attempt to check their working out. In most cases, though, there was little written evidence indicating whether or not the learners had checked their calculations. The interviews offered an opportunity to explore with the learners whether they checked their calculations and, if they did check their work, how they went about it. The findings revealed that the learners checked their calculations in different ways and at different times during the workbook task. Furthermore, the desire to check their work sometimes influenced their initial choice of strategy too.

***How learners checked their calculations*.**

Altogether 15 of the learners (6 boys, 9 girls) checked all of their calculations by re-applying their initial choice of strategy whilst seven of the learners (6 boys, 1 girl) adopted multiple approaches towards checking their work and one girl estimated her answers. The mean accuracy score for the 15 learners who relied exclusively on reapplying their initial strategy (M = 10.6, SD = 2.4) was lower than the accuracy score for the eight learners who adopted multiple approaches towards checking their work or estimated their answers (M = 12.5, SD = 2.6).

The most frequently selected checking procedure was re-applying their initial choice of strategy. Indeed, 22 out of the 23 learners (12 boys, 10 girls) checked at least some of their calculations using that approach. Ben’s response was typical, “I check to see if I've got the numbers in the right spots and they are the right numbers, and then I go over and over it to see if I've got the answer right.”

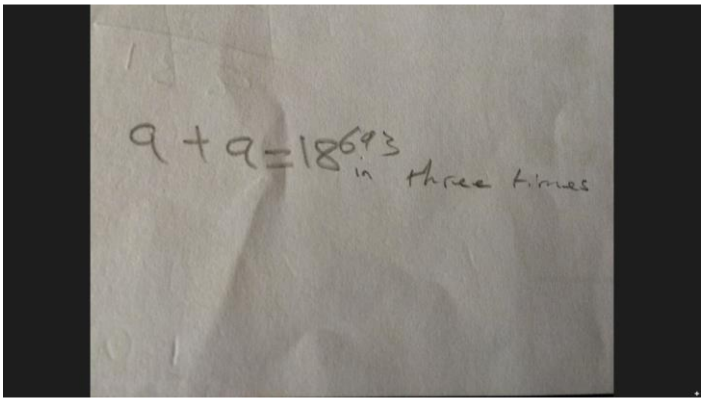
Eleven of the learners who checked their answers by re-applying their initial strategy (7 boys, 4 girls) admitted that their strategy selection was driven by the need to check their work. Those 11 learners consisted of nine higher attaining learners (5 boys, 4 girls) and two of the low confidence boys, and their mean accuracy score of 11.6 (SD = 2.3) was slightly higher than the mean accuracy score for other interviewees. Although those 11 learners often admitted that they knew multiple ways to complete a calculation, they opted for a calculation strategy which they felt they could check. For example, Edward dismissed using the partitioning approach for a particular calculation because he felt that it was “harder to check.” Several of those 11 learners were reluctant to use a mental approach. Charlie rejected working mentally because, “I wanted to write it out because I wanted to check that it is right.” Likewise, Harry reflected, “Well, I could do it in my head [832-299] but I just do [a written approach] to check.” Helen also favoured a written method over a mental one, “Because if I did it in my head I'd always want to check it afterwards." None of those learners suggested using jottings as an aide-memoire for keeping track of a calculation as well as supporting them to check their workings.

Intriguingly, two of the low confidence boys described an apparently inefficient way of checking their calculations; yet they achieved impressive accuracy scores. "Well, basically, I cleared it. I got rid of the answer and I went through the whole thing again," explained Kevin. Nevertheless, despite rubbing out their answers each time and starting again to check them, both of those boys achieved 13 correct answers which were the joint highest scores amongst the low confidence interviewees as well as higher than the mean score for the 23 interviewees.

The seven learners who adopted more than one checking procedure (6 boys, 1 girl) chose from a variety of divisibility rules, using inverses and using their number sense to check their answers. However, none of those learners checked their answers using another strategy within the same operation. The six higher attaining learners who chose more than one checking procedure (5 boys, 1 girl) achieved a mean accuracy score of 14.2 (*SD* = 1.8) compared with a mean accuracy score of 11.4 (*SD* = 1.5) for the other higher attaining learners. The low confidence boy only achieved 9 correct answers, which was lower than the mean accuracy score for the other low confidence learners (*M* = 11.5, *SD* = 1.4).

Divisibility rules.

Three of the higher attaining learners (Alex, Ben and Charlie) applied their divisibility rules to check their answers for 693 ÷3. Alex commented, "I knew it wasn't going to be a decimal number because 6 + 9 + 3 = 15 + 3 =18 which means 3 goes into 18 which means 3 goes into it." His working out is presented in Figure 16. None of the higher attaining girls, nor any of the low confidence learners, discussed using divisibility rules during their interviews.



`

Figure 16. Alex checked his calculation by using his knowledge of divisibility rules.

Inverses.

Two of the boys (one higher attaining, one low confidence) checked their division calculation 517 ÷ 19 by performing the inverse operation of multiplication. Both Edward (see Figure 17) and Fred chose the grid method to check their answers for 517 ÷ 19. None of the girls discussed using inverses to check their answers.

Number sense.

Two of the low confidence boys drew upon their number sense to check their work. Both learners re-checked calculations when they realised that their initial answers were incorrect. When Mark's first attempt at calculating 401 ÷ 25 resulted in the answer 723 he quickly realised his error, "I was like…that isn't right!" Likewise, when Fred reflected on his initial attempt at calculating 35 x 99 he admitted, "I knew it wasn't right…it just couldn't be right…" Neither boy either estimated his answers or checked them using a mental or written method; it seemed that their number sense resulted in the realisation that they had made a mistake.

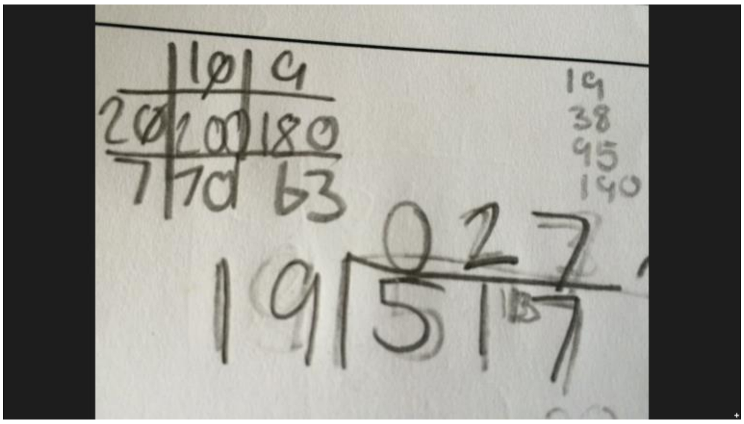


Figure 17. Edward checked his answer for 517 ÷ 19 by using the inverse operation of multiplication.

Three of the low confidence learners (1 boy, 2 girls) could not suggest any multiple strategies without prompting. "I can't think of another method I've seen," admitted Milly. When Harry was asked to name alternative addition strategies, he mistakenly suggested a division strategy, "I know the bus stop…" In contrast, all of the higher attaining learners could suggest alternative strategies for working out addition, subtraction and multiplication calculations but not necessarily for division.

Only one of the 23 learners did not check any calculations by repeating either their initial strategy or applying another calculation strategy. Instead Lucy, a low confidence learner, stated that she checked all of her calculations by estimating each answer before completing her calculation and later comparing her actual answer with her initial estimate. Her total score of eleven correct calculations was the joint top score amongst the low confidence girls.

When learners checked their calculations.

As well as checking their answers in different ways, the learners checked their answers at different points during the workbook task. Eighteen of the learners (8 boys, 10 girls) checked each of their answers before moving on to the next calculation whereas the other five learners (4 boys, 1 girl) delayed checking their answers until they had completed their workbooks.

The four boys who checked their answers at the end of the task appeared to prioritise completing their workbooks over ensuring that their work was accurate. Two of those boys, both higher attaining learners, appeared confident that they would still have time to check their work. Charlie noted, "I try to get it done as fast as I can and then check through it." Liam echoed his comments, "I just want to go through them and then check as many as I can." Nevertheless, those two boys achieved the lowest scores out of all the higher attaining boys. Only one of the girls left checking her workbook until she had completed all of the calculations. Helen, another higher attaining learner, appeared less confident than those two boys when she explained that she would check her calculations at the end, "If there was time." Helen achieved the second lowest score amongst the higher attaining girls.

The other two boys who left checking their answer until they had completed their workbooks were both low confidence learners. Like Helen, Gary appeared unsure about his ability to complete the workbook within the allocated time, "Well what I normally do is I complete the paper then I go back and check them so at least I've managed to do all the questions." Fred took a different approach by prioritising speed, "I don't know how to explain…I just wanted to finish it and get it over with." Both of those boys achieved the joint highest scores amongst the low confidence interviewees.

Two of the boys prioritised completing their workbooks before checking their answers. Low confidence learner Fred admitted, "I just want to go through them and then check as many as I can." Later in his interview he admitted, "I rush, I rush…my mum tells me!" Meanwhile, higher attaining learner Charlie also prioritised finishing his workbook, "I try and get it done as fast as I can and [then] check it though." Their mean accuracy score (M = 11.5, SD = 2.1) was lower than the mean accuracy score of the other boys (M = 11.7, SD = 2.4).

Role of adults.

Four of the interviewees (3 boys, 1 girl) also revealed details about the role of adults in developing their calculation accuracy. They included two learners (1 boy, 1 girl) who had out-of-school tutors to support their mathematics. Higher attaining learner Liam had a tutor who revised with him the same calculation strategies that were taught in his school. For example, when Liam's school introduced the formal algorithm for subtraction, "The next day my tutor went through it with me and then I knew it." Low confidence learner Milly reported that her tutor did not always follow her preferred strategies because he focused solely on formal algorithms. She spoke positively about her tutor's role when learning the formal algorithm for addition, but she struggled with long multiplication whereas Milly preferred the grid method, “I can’t really explain it because I don’t really understand the way he does it.” She seemed more confident about long division, “He taught me how to do it and now I’m starting to figure out how to do it.”

Meanwhile, two boys who prioritised completing their workbooks before checking their answers recalled how different adults had drawn attention to their errors. Charlie, the lowest scoring higher attaining boy among the interviewees, noted that his teacher cautioned him about his rounding errors. When discussing his counting up approach for correctly calculating 500 – 76, Charlie admitted that he also knew how to work it out mentally but was wary of making a mistake with his number bonds. His teacher had made him aware that he tended to make rounding errors, leading to incorrect answers, “I always get into the trap…Mrs X says…” Fred, who achieved thirteen correct answers, explained that his mum warned him that his tendency work quickly could lead to avoidable errors, “When I look at a mistake, I get it!”

Integrating the findings for RQ2.

Although the quantitative results revealed that confidence and prior attainment were both significant predictors of calculation accuracy, a large amount of the variance remained unexplained in the regression model. The qualitative findings shed further light on those quantitative results by revealing that practice, checking procedures and the role of adults were three further variables affecting calculation accuracy.

The quantitative results showed that the learners were more accurate adding and subtracting rather than multiplying or dividing. The subsequent qualitative analysis revealed that just under half of the learners lacked practice with their multiplication or division strategies. The higher attaining learners tended to work more accurately than the other learners. The interviews subsequently revealed that only two of the higher attaining learners felt confident applying all of their chosen strategies, and the data showed that their mean score was higher than the mean score for the other learners. Also, more of the higher attaining learners claimed to practice their chosen multiplication and division strategies than the other learners.

The interviews also showed that the learners checked their work at different times during their task and in different ways too. The mean score for learners who checked their work using multiple approaches was higher than those learners who maintained a single approach.
Also, most of the learners preferred to check their answers as they went along rather than wait until they had completed all of their calculations. Their mean score was higher than the mean score for the two learners who checked their work at the end.

The quantitative data for RQ1 indicated that the girls were more likely to choose formal algorithms than the boys. The subsequent interviews showed that most of the girls preferred to use the same strategy for calculating their answers and checking their work whereas more of the boys adopted different checking techniques such as applying their knowledge of divisibility rules. Those boys appeared to have unwittingly generated more opportunities to practise a wider range of strategies than the girls.

The quantitative results also showed that the low confidence learners tended to work less accurately than the other learners. The interviews showed all of the low confidence boys appeared to lack practice with either their multiplication or division strategies.

Only two of the learners discussed out-of-school tutors, one of those learners felt that his tutor supported his learning by focusing on the same strategies taught in class whereas the other learner sometimes wanted to practice different strategies.

Phase 2 Findings for RQ3

The quantitative analysis (*N* = 590) revealed that formal algorithms were the most frequently selected strategy for each of the 16 calculations, but other strategies were significantly more accurate than formal algorithms for four of those 16 calculations (702 – 695, 27 x 63, 401 ÷ 25 and 517 ÷ 19). The interviewees' workbooks revealed that formal algorithms were their most frequently selected strategy for those three calculations (Figure 18).

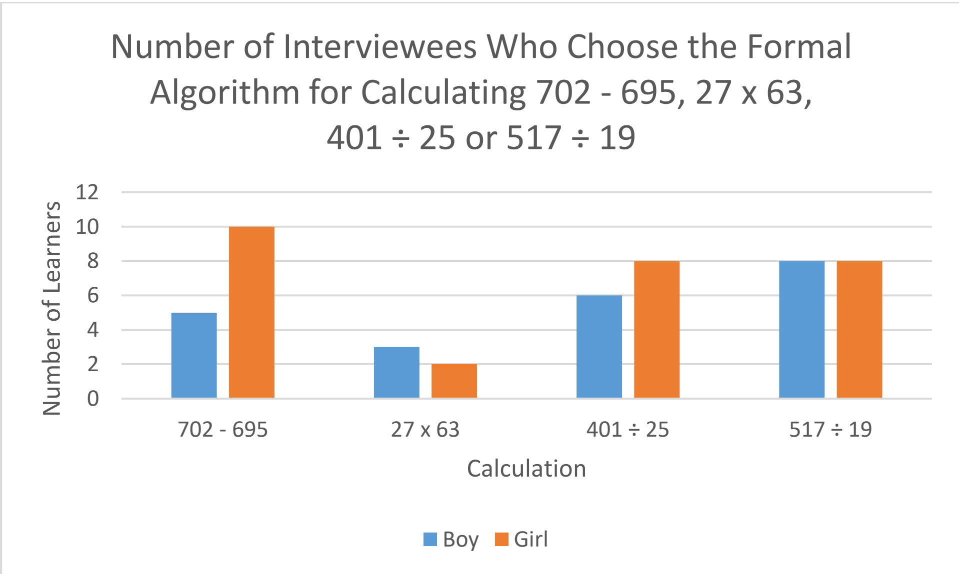


Figure 18. Number of interviewees who chose formal algorithms for the calculations 702 – 695, 27 x 63, 401 ÷ 25 or 517 ÷ 19 (*n* = 23).

Two of those four calculations were included in the research design due to their deemed suitability for using a strategy other than the formal algorithm (702 – 695 and 401 ÷ 25). Although the near difference between 702 and 695 could have been calculated using a counting-up strategy, more than half of the learners (5 boys, 10 girls) chose the formal algorithm; all of those boys reached the correct answer but two of the girls made errors, one of those girls transposed the digits and the other one mistakenly added the numbers together.

Similarly, the calculation 401 ÷ 25 was deemed suitable for using a strategy other than a formal algorithm, such as counting-up in steps of 25. Nevertheless, more than half of the learners (6 boys, 8 girls) chose the formal algorithm but only six of them reached the correct answer (3 boys, 3 girls). The calculations 27 x 63 and 517 ÷ 19 appeared more suitable for using a formal algorithm than the other two calculations. Nevertheless, the formal algorithm was not the most accurate approach in this study for either of those calculations because more of the learners were successful using the grid method and chunking down procedure respectively. Among the interviewees, five of the learners (2 girls, 3 boys) chose the formal algorithm for calculating 27 x 63 and two of them (1 girl, 1 boy) reached the right answer. For 517 ÷ 19, 16 of the learners (8 boys, 8 girls) chose the formal algorithm but only three of them (all boys) reached the correct answer.

Reasons for choosing the formal algorithm for 702 – 695, 27 x 63, 401 ÷ 25 and 517 ÷19.

During their interviews, the learners were asked to justify their strategy choices for each of the 16 calculations. The responses from the learners who chose a formal algorithm to calculate 702 – 695, 27 x 63, 401 ÷ 25 or 517 ÷ 19 were collated and analysed for recurrent themes among the responses (Figure 19).

Efficiency was the most frequently cited reason for choosing a formal algorithm for the four calculations. Reflecting on her choice of the formal algorithm to calculate 702 – 695, Gabby noted, “It’s a very simple way I learnt.” Six of the learners stated that the formal algorithm was their only known division strategy. “It’s just what I know,” claimed Charlie. Practice was also important. For example, two of their girls highlighted their division practice using the formal algorithm. “I’ve been doing it a while,” noted Daisy. “I’ve had lots of practice,” Lucy reflected, “We learnt the bus stop in Year Four.”

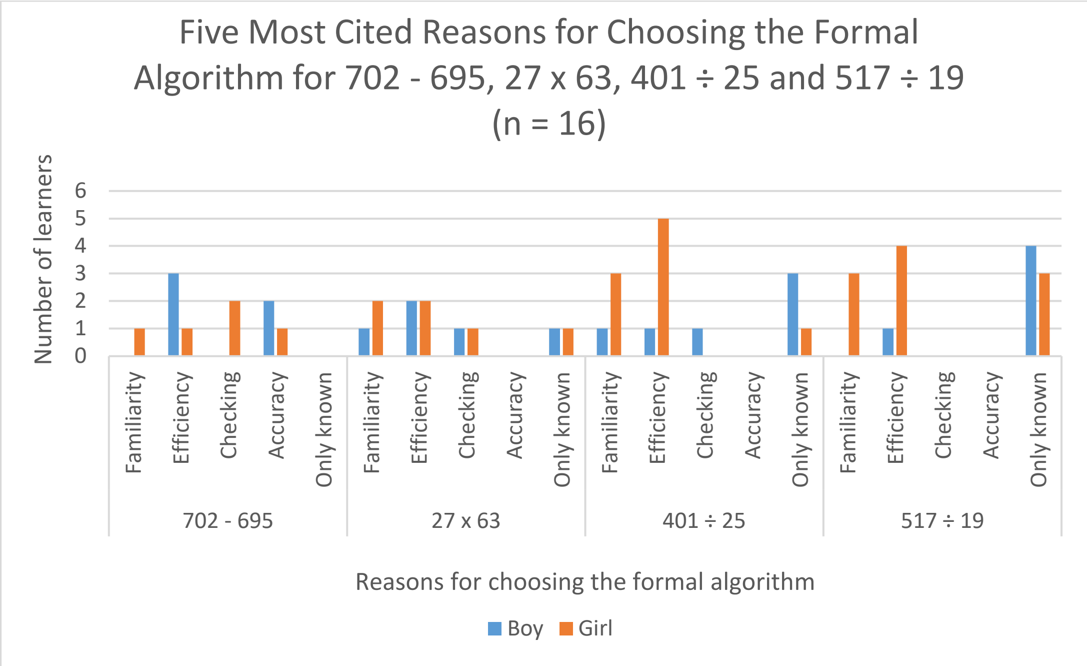


Figure 19. Five most cited reasons for choosing the formal algorithm for 702 – 695, 27 x 63, 401 ÷ 25 or 517 ÷ 19 (n = 16).

Integrating the findings for RQ3.

Although the quantitative results revealed that the formal algorithms were the most frequently selected strategies across the 16 calculations, other strategies were significantly more accurate for four of those calculations (702 – 695, 27 x 63, 401 ÷ 25 and 517 ÷ 19). The qualitative aspect of the study explored the reasons given by the learners who had chosen formal algorithms for those calculations. Their most cited reasons addressed the perceived efficiency of the formal algorithm as well as their time practising with that approach. However, a quarter of the interviewees who chose the formal algorithm for division admitted that it was their only known division strategy.

So far, this chapter has focused on the questions arising from the quantitative analysis relating to the first three of the four RQs. The second part of this chapter focuses on RQ4 which addressed calculation efficiency and conceptual understanding.

Phase 2 Findings for RQ4

RQ4: To what extent do calculation efficiency and understanding vary by gender, prior attainment and confidence?

In this section, the findings from the qualitative analysis relating to both calculation efficiency and understanding will be presented, beginning with calculation efficiency.

Efficiency.

Calculation efficiency was much more challenging to analyse than calculation accuracy or use of the formal algorithm. Whilst an individual learner's calculation accuracy was analysed according to their total number of correct answers, and a similar approach was adopted for use of formal algorithms, analysing their calculation efficiency demanded a different approach. Although an individual learner's selected strategy might have appeared inefficient, the literature review drew attention to learners developing their calculation skills at different rates. Hence, an efficient strategy for one learner might have seemed inefficient for another one. By defining an efficient strategy as "one that the student can carry out easily" (Russell, 2000, p. 5), the analysis attempted to take into account those developmental differences. Also, learners might develop their strategies across the four operations at different rates. For example, a learner might subtract efficiently but find multiplying much more challenging. Thus the qualitative phase of the study offered an opportunity to explore the perspective of the learners themselves regarding calculation efficiency by analysing their responses which referred to the perceived easiness or simplicity of their chosen strategy according to their gender, prior attainment and confidence levels.

Twenty-two out of the 23 interviewees (11 boys, 11 girls) justified choosing at least one of their strategies by referring to its perceived efficiency. Those 23 learners included ten higher attaining learners (5 boys, 5 girls) and 12 low confidence learners (6 boys, 6 girls). Overall the interviewees made a total of 59 comments justifying the choice of a calculation strategy by referring to its perceived efficiency. The boys made a total of 28 comments compared to 31 comments by the girls There were 31 comments from the higher attaining learners (18 from boys, 13 from girls) and 28 comments from the low confidence learners (14 from boys, 14 from girls).

The findings became more revealing when the analysis compared the number of comments referring to the perceived efficiency of formal algorithms with those referring to other strategies. For the higher attaining learners, 20 out of their 31 comments referred to the perceived efficiency of the formal algorithm. The proportion was even higher among the low confidence learners, 21 out of the 28 comments referred to the perceived efficiency of the formal algorithm. Focusing on gender, the proportion was slightly higher for the girls than the boys; 19 out of the 28 comments by the boys referred to the perceived efficiency of formal algorithm compared with 22 out of the 31 comments by the girls.

***Formal algorithms*.**

Almost all of the learners (9 boys, 11 girls) justified at least once choosing the formal algorithm for addition by referring to its perceived efficiency. All of the girls and the low confidence boys chose the formal algorithm for all of their four addition calculations, and all of those girls and five of the low confidence boys justified their decisions by referring to the perceived efficiency of that strategy.

Amy’s comment was typical, “I find it’s the easiest way to do it.” Although some of the higher attaining boys varied their addition strategies, only three of those boys justified their decisions by referring to its perceived efficiency.

Just under half of the learners (5 boys, 5 girls) justified choosing the formal algorithm for subtraction by referring to its perceived efficiency which was half of the number who justified its use for addition. Although all of the girls chose the formal algorithm for at least one of their subtraction calculations, and five of them chose it for all of their subtraction calculations, just five of the girls (2 higher attaining, 3 low confidence) justified their decision based on its perceived efficiency. Helen felt that the subtraction formal algorithm was “quite easy” whereas Milly noted that “It’s easier to write down quickly.” Only four out of the eight boys who chose the formal algorithm for any of their subtraction calculations referred to its perceived efficiency (1 higher attaining, 3 low confidence). Unlike the girls, they all made connections between their strategy choices and the numbers in the individual calculations. For example, James explained his reason for choosing the formal algorithm to calculate 702 – 695, “I think I find the sum, like, easier when there’s a zero at the top.” Similarly, Alex selected the formal algorithm to calculate 382 – 199 after reflecting on the numbers involved, “Like, it will go into a different hundred so I find it easier to write up on paper.” Just one of the learners commented on the perceived efficiency of the formal algorithm for multiplication but it should be remembered that only a quarter of the learners attempted the formal algorithm for any of their four multiplication calculations. James explained why he chose the formal algorithm rather than the grid method, “I don’t find it as easy as that [formal algorithm].”

Nineteen of the learners chose the formal algorithm for division for at least one of their division calculations, but only six of those learners (1 boy, 5 girls) commented on its perceived efficiency. Four of those were higher attaining (1 boy, 3 girls) and two were low confidence (both girls). Edward was the only boy who justified choosing the formal algorithm for his division calculations based on its perceived efficiency. Indeed, he was a very enthusiastic advocate of that approach. Nevertheless, more of the girls commented on the perceived efficiency of the formal algorithm for division. “It’s a very simple way I learnt at school,” noted Gabby.

Other strategies.

The only learners who chose any strategies other than the formal algorithm for any of their addition calculations were three boys (2 higher attaining, 1 low confidence). Each of them justified their decisions by referring to the perceived efficiency of their chosen approaches. Low confidence David chose partitioning for each of his addition calculations (Figure 20). Although his approach might be interpreted as an inefficient strategy, his workbook revealed that he made few jottings and he correctly answered each of his four addition calculations.

Edward adopted a different approach towards his addition strategies. He switched strategies from using the formal algorithm for three of his addition calculations to choosing rounding to calculate 299 + 532, "I thought it would be an easier way. Sometimes, for different sums I use different methods…I thought it would be an easier way."

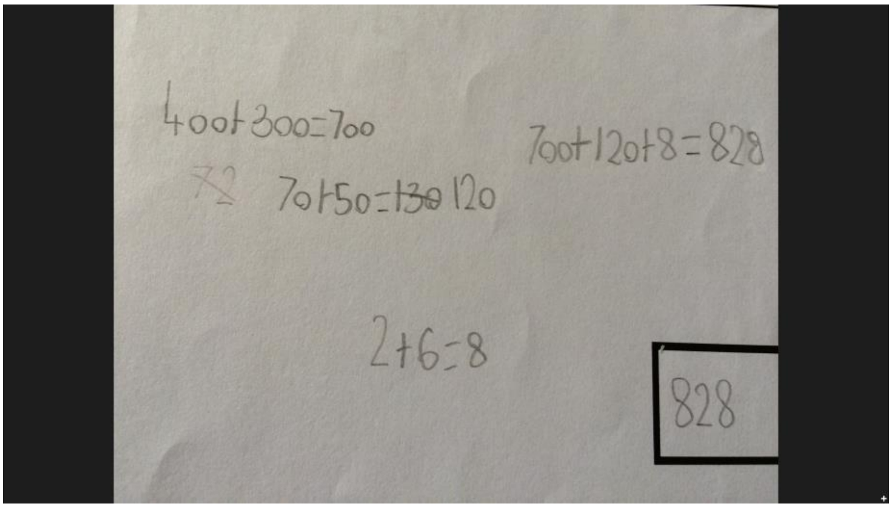


Figure 20. David performed his addition calculations by partitioning.

Just four of the learners (2 girls, 2 boys) referred to the perceived efficiency of their subtraction strategies which were not the formal algorithm. Three of those learners (2 boys, 1 girl) were higher attaining learners who switched from the formal algorithm to another strategy for at least one of their calculations. Higher attaining learners Beth and Edward preferred the efficiency of a counting-up approach for calculating 500 – 76 and 702 – 695 respectively, whereas Ben drew upon his number bonds to calculate 500 - 76. In contrast, low confidence learner Lucy decided to use counting back to calculate 500 – 76 but Kevin chose partitioning for that particular calculation, "I find in some situations it's easier to partition than use the column [formal algorithm]."

The grid method was the most frequently selected multiplication strategy by the interviewees, it was chosen at least once by 18 of the 23 learners. However, only five of those 18 learners (2 boys, 3 girls) commented on its perceived efficiency. Two of those learners (1 boy, 1 girl) were higher attaining learners who chose the grid method for all of their four multiplication calculations. "I find the long multiplication more complicated," admitted Daisy. Similarly, Alex noted, "I find that way [the grid] easier." The three low confidence learners (1 boy, 2 girls) who also chose the grid method for all of their multiplication calculations noted its perceived efficiency. "It's a lot easier," noted Gary.

Altogether five of the learners (4 boys, 1 girl) commented on the perceived efficiency of their division strategies which differed from the formal algorithm. Those children included two of the higher attaining learners, Daisy and Liam, who calculated 401 ÷ 25 by counting-up. "That's the slowest but the easiest way," reflected Liam. None of the low confidence girls attempted any strategy other than the formal algorithm for their division calculations but David felt that it was easier to use counting to calculate 401 ÷ 25 rather than apply the formal algorithm, "It fitted, the 25 would be easier to go into the hundred so you do it like that." Likewise, Kevin counted-up too "It sounds a bit simple but I can do it." In contrast, Fred felt that it was easier to draw upon his knowledge of multiples of ten to calculate 480 ÷ 20, than apply the formal algorithm.

Conceptual understanding.

The literature review highlighted different interpretations of the word 'understanding.' For example, a learner might think that they 'understood' a strategy if they simply 'got it.' In other words, they might interpret understanding as the ability to follow their teacher's example and later apply the same approach themselves to a similar calculation. However the literature showed that the level of understanding required for achieving fluency arguably went much deeper, beyond merely following a taught procedure. In particular, Kilpatrick et al. (2001, p. 118) suggested that if learners fully understood their approach, they would be less likely to recall it incorrectly. They argued that learners demonstrating such 'conceptual understanding' should identify their errors, make connections and possess a comprehensive knowledge of the four calculation operations. Drawing upon their four key aspects of conceptual understanding, the extent of an individual learner's conceptual understanding was analysed by confirming whether or not they:

- correctly applied their chosen approach
- identified any errors
- made connections between the operations
- knew alternative strategies

However, it seemed unrealistic to expect the Y6 learners to demonstrate all four of those aspects for each of their calculations. For example, if they had made an error using their chosen multiplication approach but spotted their mistake during their interview, then it seemed reasonable to accept their conceptual understanding of that approach if they also satisfied the other criteria addressing making connections and knowing alternative strategies. If there was sufficient evidence that an individual learner satisfied at least three out of the four requirements, then it was accepted that they had demonstrated an acceptable level of conceptual understanding for that particular operation. However, just three out of the 23 interviewees (2 boys, 1 girl) demonstrated sufficient conceptual understanding to satisfy those criteria for all of the four operations but several of the learners did demonstrate their conceptual understanding for some, if not all, of those operations.

***Addition*.**

There were 22 of the learners (11 boys, 11 girls) who satisfied at least three of the four criteria. which was the highest number of learners demonstrating their conceptual understanding for any of the four operations. Only one of the low confidence boys failed to reach that standard; Kevin correctly applied his preferred formal algorithm strategy but struggled naming or describing alternative strategies until prompted, "I'm guessing you could do the partition or number line." The other 22 learners were usually able to correctly apply their preferred method and suggest an alternative strategy too, although they did not necessarily choose to use that alternative strategy for a variety of reasons.

***Subtraction*.**

Altogether 20 of the learners (11 boys, 9 girls) satisfied at least three of the criteria for demonstrating their conceptual understanding of subtraction. Although those figures included all of the higher attaining learners, only nine of the low confidence learners (5 boys, 4 girls) also met at least three of the criteria. Higher attaining learner Alex was one of the few learners who gave evidence about making connections between the four operations. For example, he explained how he checked his answer to the subtraction calculation 500 – 76 by using the inverse operation. Again, Kevin was the only low confidence boy who did not satisfy at least three of the criteria for conceptual understanding. He incorrectly applied the formal algorithm to one calculation, without noticing his error, and struggled to suggest alternative approaches except making a vague reference about partitioning.

Two of the low confidence girls also failed to satisfy at least three of the criteria for demonstrating their conceptual understanding of subtraction. Both Kate and Milly struggled to reach the correct answers or identify their mistakes. Indeed, Kate gave incorrect answers for each of her four subtraction calculations. As discussed earlier, she mistakenly added rather than subtracted her numbers and did not realise her error without prompting during her interview. However, Kate was fully aware of her weaker subtraction skills in comparison with her knowledge of addition, "[When] we are doing about add and take aways, sometimes I get them muddled up." Although Milly did not confuse her addition and subtraction calculations, she incorrectly calculated two of her subtraction calculations and could not identify her errors without support.

It should be noted that there was only a slight difference in the numbers of learners satisfying at least three of the criteria for demonstrating their conceptual understanding of multiplication and division, but the figures for both operations were much lower than those for addition and subtraction.

Multiplication.

Less than half of the learners (7 boys, 4 girls) satisfied at least three of the criteria for demonstrating their conceptual understanding of multiplication. The grid method was the most frequently selected multiplication strategy by the interviewees, but it attracted the most mistakes too. In most cases, the mistakes highlighted a lack of understanding about place value and the effects of multiplication on numbers.

Beginning with the higher attaining learners, Beth made errors which she did not identify without prompting. For example, she suggested that the product of 20 x 46 was 966 rather than 920. She also miscalculated the product of 27 x 63 as 621 rather than 1701 (Figure 21).

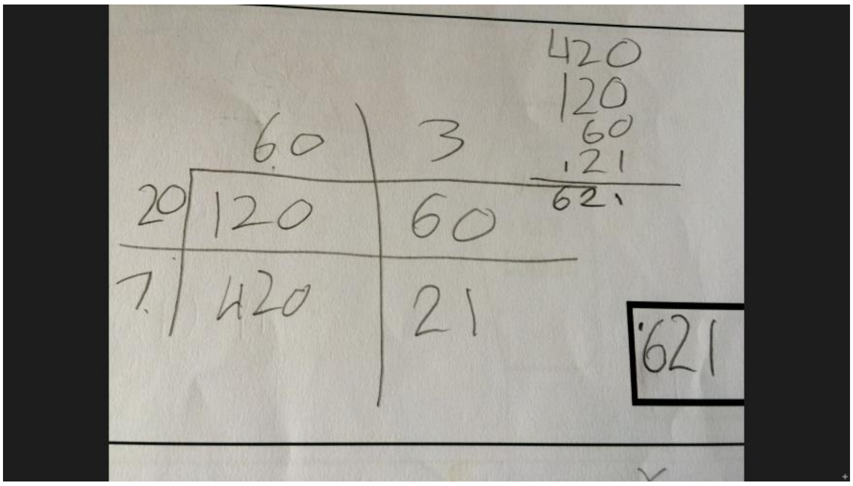


Figure 21. Beth miscalculated the partial products of 27 x 63.

Another higher attaining learner, Claire, could correctly calculate her partial products but she struggled adding them together and failed to recognise her mistakes until prompted to do so during her interview. During his interview, Liam reflected that he preferred using the grid method but struggled applying it to calculations involving three-digit numbers. Following up his comments with the wider sample (N = 590) revealed that 89 learners, just over a seventh of the sample, chose different strategies for calculating 27 x 63 and 568 x 34. This was surprising since both calculations were deemed suitable for the formal algorithm during the design phase. Over a quarter of those 89 learners who changed strategies abandoned the grid method when faced with the three-digit multiplication calculation, and nine of those learners switched to the formal algorithm. Interestingly, the changing of strategies worked both ways when faced with a three-digit multiplication calculation; around the same number of learners switched from using the formal algorithm to the grid method.

Only four of the low confidence learners (3 boys, 1 girl) satisfied at least three of the criteria for demonstrating their conceptual understanding of multiplication. The mistakes among the low confidence learners highlighted a lack of understanding about the effects of multiplication on numbers. For example, David gave incorrect answers for all four of his multiplication calculations and his answers highlighted his lack of understanding about place value such as incorrectly calculating the product of 20 x 46 as 9,200 rather than 920.

Fred also made several errors too. He chose the grid method for three of this calculations but preferred using the formal algorithm to calculate 35 x 99. His answers were much closer to the actual answers than David's efforts, but Fred did not consistently apply his procedures correctly and failed to identify his errors without prompting. Meanwhile, Harry, Kevin and Liam all incorrectly calculated the partial products for 20 x 46 using the grid method. Five of the low confidence girls, all using the grid method, also struggled with the same calculation.

Division.

Although twice as many boys as girls (6 boys, 3 girls) satisfied at least three of the criteria for demonstrating their conceptual understanding of division, it was the operation with the fewest number of learners meeting the criteria. Eight of those learners were higher attaining learners (5 boys, 3 girls) and the other learner was a low confidence boy. None of the low confidence girls satisfied the criteria.

Charlie was the only one of the higher attaining boys who failed to satisfy at least three of the criteria for demonstrating his conceptual understanding of division. He chose the formal algorithm for division but struggled to name any alternative approaches. Moreover, two of his answers were wildly inaccurate too. For the calculation 517 ÷ 19, he gave the answer as 210 remainder 7/19. Similarly, he calculated that 401 ÷ 25 was 110 remainder ¼. Higher attaining girls Amy and Beth also struggled with their division skills. Amy attempted the formal algorithm for each of her four calculations but only achieved a correct answer for the calculation 693 ÷ 3. Beth did attempt two different strategies for her division calculations, the formal algorithm and counting up, but she made errors too.

Only one of the low confidence learners satisfied at least three of the criteria to demonstrate his conceptual understanding of division. Surprisingly, Kevin - who struggled with the addition and subtraction calculations - correctly applied a variety of division approaches. However, all of the other low confidence learners struggled with their conceptual understanding of division. Indeed, James simply did not record any answers whilst Mark admitted guessing his answers. Most of the low confidence learners struggled to suggest alternative division approaches and several gave responses which highlighted their lack of understanding. For example, Gabby incorrectly suggested that the answer for 401 ÷ 25 was 200 whilst Kate suggested that it was 20,000.

The analysis also revealed a misunderstanding regarding the range of applications for the formal algorithm. Three learners (2 boys, 1 girl) believed that the formal algorithm for division was exclusively reserved for single digit divisors with multi-digit dividends, such as 693 ÷ 3. Emily explained, "The standard algorithm [formal algorithm] is for ones and big numbers." David admitted choosing the formal algorithm for calculating 693 ÷ 3, "because it's a tiny number and a big one" but rejecting it for other division calculations which did not satisfy those criteria. James, who did not record an answer for any of the four division calculations, admitted he struggled when faced with the first division calculation in his workbook (401 ÷ 25) because of the number of digits, "I like doing it [formal algorithm] with lower numbers." Likewise he also struggled to cope with the second division calculation (480 ÷ 20) which also had a two-digit divisor. By the time James reached the third division calculation which only had a single digit divisor, 693 ÷ 3, he had already decided to abandon the division calculations. Hence James, who correctly answered all of the other twelve calculations in his workbook, scored nothing for his four division calculations. Further analysis of the wider sample (N = 590) revealed that James, Emily and David were not the only learners who may have believed that the formal algorithm for division was restricted towards single digit divisors with multi-digit dividends. Around a tenth of the learners who selected the formal algorithm to calculate 693 ÷ 3 did not choose it for any of their other three division calculations.

Integrating the findings for RQ4.

The qualitative analysis offered an insight into the perceived efficiency and conceptual understanding regarding the chosen calculation strategies in respect of gender, prior attainment and confidence. The vast majority of the learners justified at least one of their calculation choices by referring to its perceived efficiency. However, the girls made more comments about the perceived efficiency of the formal algorithms whereas the boys focused on other strategies. For both genders, most of their efficiency comments referred to the addition calculations, and the least of their comments referred to the multiplication calculations. Most of the higher attaining learners commented on the perceived efficiency of their calculation strategies, and there were twice as many comments about the perceived efficiency of formal algorithms than other strategies. Again, most of the comments addressed the perceived efficiency of addition approaches and the fewest comments referred towards multiplication. All of the low confidence learners commented on the perceived efficiency of their chosen strategies. There were around three times as many comments referring to formal algorithms compared to other strategies.

Only three out of the 23 learners, all higher attaining learners, demonstrated their conceptual understanding of all of the four operations. Addition was the only one of the four operations where more girls than boys demonstrated their conceptual understanding. All of the higher attaining learners demonstrated their conceptual understanding of both addition and subtraction, but the numbers were much lower for multiplication and division. Only the higher attaining boys drew upon their divisibility rules to check their answers. Most of the low confidence learners demonstrated their conceptual understanding of addition and subtraction, but the numbers were much lower for multiplication and division.

The learner interviews revealed a lack of understanding relating to the grid method and the formal algorithm for division. Some of the learners struggled to extend the grid method to three-digit numbers. Moreover, some of the learners believed that the formal algorithm for division was only suitable for calculations with single-digit divisors.

Summary

The regression models arising from the quantitative analysis only partially addressed the RQs. However, the explanatory sequential nature of this study enabled further exploration of those results. The qualitative findings for RQ1 and RQ2 revealed further variables affecting use of the formal algorithm and calculation accuracy respectively. They also addressed the reasons behind the calculation choices relating to RQ3. Furthermore, the qualitative findings for RQ4 indicated a number of variables relating to calculation efficiency and conceptual understanding, including misconceptions about the range of calculations suitable for some strategies. The next chapter will consider the meaning and importance of these findings as well as assess their contribution to the field of mathematics education.

Chapter Six: Discussion and Conclusions

Introduction

In our increasingly automated world, nurturing future generations of learners with the ability and inclination to solve non-routine problems is perhaps more important than ever before. Since younger learners tend to follow a mathematics curriculum dominated by number and calculation, their early teaching could focus on nurturing their calculation fluency by encouraging them to work flexibly and choose the most efficient strategy for each of their calculations rather than adopting a ‘one size fits all approach.’ However, the literature review revealed heated debates regarding the relative merits of focusing on conceptual understanding or procedural fluency (Klein, 2007). For the purposes of this thesis, informed by the findings from the literature review, learners demonstrating their calculation fluency worked accurately, efficiently, flexibly and with conceptual understanding. Nevertheless, the recent reforms (DfE, 2013a) which prioritised the use of formal algorithms over and above other calculation strategies appeared to restrict the opportunities for younger learners to develop their flexibility and choose the most efficient strategy for individual calculations. Moreover, the literature showed that formal algorithms were not necessarily the most accurate strategies for individual calculations (Borthwick & Harcourt-Heath, 2007).

Gender, confidence and prior attainment were key factors for this study. The literature indicated gender differences in calculation strategies (including Carr & Davis, 2001; Fennema et al., 1998) and concerns regarding the mathematical confidence of girls (including Nunes et al., 2009). Confidence was also regarded as a key issue regarding post-16 mathematical study (Brown et al., 2008; Heilbronner, 2013; Pampaka et al., 2011). The literature highlighted gender imbalances favouring boys in mathematical performance (Gill, 2015; OECD, 2014) and the proportion of students pursuing mathematical careers (McWhinnie & Fox, 2013). The theoretical model proposed by Villalobos (2009) predicted that girls who thrived under a curriculum focusing on number and calculation at primary level might struggle when they progressed towards a curriculum focusing on problem-solving in their secondary schools. Although traditional pencil-and-paper approaches such as using log tables had fallen into disuse with the growth in access to calculators, the literature raised concerns regarding the relevance of continuing to teach written calculation strategies (Keiser, 2012; Kilpatrick et al., 2001; Plunkett, 1979; Wolfram, 2014).

Drawing together the key concerns outlined above, this thesis addressed the overarching question ‘Does calculation fluency vary by gender, confidence level and prior attainment?’ with the first cohort of Y6 learners studying under the DfE’s (2013a) reforms. The mixed methods study was conducted in two phases: the initial quantitative phase surveyed the calculation strategies and confidence levels of a sample of Y6 learners (N = 590), and the subsequent qualitative phase interviewed a purposive sub-sample of those learners (n = 23). The findings, which will be elaborated on further in this chapter, indicated that many learners worked inflexibly, prioritising the formal algorithm over and above other strategies and regardless of the merits of individual calculations. Many learners appeared to lack practice with their strategies, especially their multiplication and division strategies, and their checking procedures relied on repeating their initial approach rather than drawing upon their range of calculation skills. Since the findings also indicated that formal algorithms were not necessarily the most accurate or efficient strategies, and that many learners lacked conceptual understanding of their strategies, the findings raised questions about the focus of the DfE's (2013a) reforms on formal algorithms.

Each of the first four sections of this chapter focuses on a single RQ, evaluating the findings considering the literature. Then the discussion continues by reflecting on the choice of an MMR design for this study, the challenges which arose during the research and an acknowledgement of the limitations of the study. This chapter concludes with a consideration of the possible implications of my findings and a personal reflection on my journey as a researcher.

Discussion of the Main Findings

RQ1: To what extent do gender, prior attainment and confidence predict use of the formal algorithm?

The literature review noted the benefits of working flexibly (including Cockcroft, 1982; Ma, 1999; Russell, 2000). Moreover, Anghileri (2000) drew attention to the close relationship between number sense and calculation, "Although children may learn some standard procedures, using number sense involves departure from these methods where the numbers warrant a different approach" (p. 127), and Gilmore and Bryant (2008) suggested that learners needed to use their methods selectively based on the merits of each individual calculation. Nevertheless, the DfE's (2013a) reforms appeared to restrict the opportunities to work flexibly during calculation sessions by prioritising formal algorithms. For the purposes of this study, considering the local context following the reforms, learners were judged to be working flexibly if they deviated from using the formal algorithm. My findings showed that formal algorithms were the most frequently selected strategies by the Y6 learners for each of the 16 calculations, even though my research design included a range of calculations across the four operations which were deemed suitable for other calculation strategies by an independent panel. This finding was in line with the findings of Torbeyns and Verschaffel (2013), who reported a tendency among 9- and 10-year-old Belgian learners who had been introduced to the formal algorithms for addition and subtraction to favour those strategies, by showing a similar tendency of English Y6 learners working across all the four operations and working with a range of calculations. Moreover, my findings showed that many learners chose formal algorithms when they were permitted to choose from their full range of calculations whereas Torbeyns and Verschaffel's (2013) choice/no choice research design necessarily restricted the strategy choices of its participants.

My quantitative findings should be regarded as indicative rather than conclusive. Gender, prior attainment and confidence were statistically significant predictors of learners using formal algorithms, and that confidence was the most significant of those predictors. The subsequent qualitative analysis revealed several other factors affecting use of formal algorithms. In particular, knowledge of alternative strategies and the SATs marking scheme also appeared to influence whether learners chose a formal algorithm. The next three paragraphs will discuss the findings for each of the three predictor variables, considering the findings from the literature review, beginning with gender.

For gender, the findings indicated that the Y6 girls were more likely to choose formal algorithms than the Y6 boys. Bearing in mind their ages, these findings were consistent with the early stages of the theoretical model proposed by Villalobos (2009) whereby young girls experienced 'strategic alignment' when they chose to follow known procedures rather than work flexibly. If the model also held for their later studies, which was beyond the scope of this study, then we might infer that those girls might struggle to adapt to the problem-solving demands of their later mathematical courses.

The findings also showed that the higher attaining learners were less likely to choose formal algorithms than the other learners. Russell (2000) argued that learners willing to adapt their approaches enjoyed a problem-solving advantage because they could change their strategy if they got stuck. My findings suggested that the higher attaining Y6 learners, who appeared more aware of alternative strategies than the other learners in the interviews, were more likely to enjoy that advantage across the four operations. Nevertheless, my findings also showed that a willingness to deviate from a specific strategy may need continual nurturing; some of the higher attaining learners expressed a reluctance to apply strategies other than formal algorithms, and others claimed that they did not recognise the opportunity to apply a different strategy. Fielker (2007) described a higher attaining teenager who chose to calculate 20 x 36 using the formal algorithm rather than perform a mental calculation; failing to recognise an opportunity to deviate from formal algorithms sometimes resulted in the Y6 learners in this study choosing an apparently less efficient, and potentially error prone, approach. For example, over 80% of the learners calculated 702 – 695 using the formal algorithm but the counting-up strategy, which was chosen by far fewer learners, was a significantly more accurate approach. We might infer that the decision-making of the Y6 learners might have reflected the high prestige afforded to formal algorithms reported in the literature (including Cockcroft, 1982; Torbeyns & Verschaffel, 2013).

Although the less confident Y6 learners were less likely to choose formal algorithms than their more confident counterparts, it should not be assumed that they were working more flexibly than the others. Their interviews revealed that some of the less confident learners had not been introduced to the formal algorithms for multiplication or division, and some of the other less confident learners lacked practice with formal algorithms. These findings also highlighted the benefits of choosing a research design incorporating the four operations since it enabled comparison of the findings between them.

Several of the interviewees also drew attention to the role of SATs in their decision-making process. They interpreted the instruction 'Show your working' as an indication that they should not deviate from formal algorithms. They did not appear to know that correct answers in SATs were rewarded with full marks regardless of their chosen strategy. Their responses offered a fascinating insight into the decision-making processes of the first cohort of Y6 learners studying under the reforms; it appeared that some learners believed that formal algorithms were the only acceptable strategy for SATs. This belief, which may have reflected the high prestige afforded to formal algorithms discussed earlier in this section, limited their opportunities for working flexibly in the workbook tasks and raised questions about their willingness to work flexibly at other times too. Their tendency to prioritise formal algorithms also resulted in their selection of strategies that the findings, explored in more depth in the following sections, indicated may have been less accurate and less efficient than alternative approaches.

RQ2: To what extent do gender, prior attainment or confidence predict calculation accuracy?

The literature review confirmed that accuracy was generally considered to be an essential aspect of calculation fluency (Gojak, 2012; NCTM, 2000; Russell, 2000). The first phase of this study addressed accuracy by considering the total number of correct answers across the set of 16 calculations for each learner. However, since their workbooks did not necessarily reveal whether learners estimated, checked their work or reflected on the reasonableness of their answers, those aspects of their calculation accuracy were addressed during the interviews.

Beginning with gender, the findings reported no significant gender differences in calculation accuracy among the Y6 learners. This was perhaps not a surprising finding; for example, the literature review included a study by US-based Fennema et al. (1998) who also reported no significant gender differences in the accuracy rates of young children performing calculations.

Moving on to prior attainment, my findings indicated that the higher attaining learners tended to work more accurately than the other learners. Again, although was perhaps not a surprising result, my research design enabled the reasons behind their success to be examined in more depth during the interviews. Practice appeared to be an important factor; the higher attaining learners who claimed to practise their chosen strategies achieved higher scores than the other learners. Moreover, since the findings from RQ1 indicated that the higher attainers were more likely to choose alternative strategies that other learners, we might infer that calculation accuracy could be increased by encouraging learners to work flexibly where appropriate.

For the confidence aspect of calculation accuracy, the literature review indicated a correlation between mathematical anxiety and calculation skills (Krinzinger et al., 2009; Thomas & Dowker, 2000). There was a gap in the literature regarding the relationship, if any, between mathematical confidence and calculation skills. My findings showed that the more confident Y6 learners tended to work more accurately than the other learners. The follow-up interviews indicated that practice appeared to be a key factor; all the low confidence learners lacked practice with either their multiplication or division strategies. One interpretation of the above findings would be to assume that learners required additional practice time. However, if Villalobos' (2009) model was correct, then focusing too strongly on procedures might increase the 'strategic alignment' of the girls and exacerbate concerns regarding their capacity to adapt to a problem-solving curriculum later in their schooling. An alternative approach might involve revisiting the level of challenge demanded by the reformed curriculum (DfE, 2013a); learners could be encouraged to develop their critical judgement (Star et al., 2016) so that they could more easily recognise suitable occasions to choose a calculator or calculate mentally, freeing up more curriculum time for developing their non-routine problem-solving skills.

Efficiency, checking and the reflecting on the reasonableness of answers.

The interviews provided insights into three other aspects of accuracy raised in the literature, namely estimating, checking and reflecting on the reasonableness of an answer. Consider estimating. Although the literature highlighted the importance of estimating (including Cockcroft, 1982; Threadgill & Sowder, 1984), there was a scarcity of recent studies exploring estimation with learners. From my findings we might infer that estimating skills were under-used by most of the learners; only one of the interviewees described estimating her answers.

Moving on to checking answers, although the literature highlighted the importance of checking (including Gojak, 2012; Keiser, 2010), there was also a scarcity of recent studies investigating this aspect of calculation accuracy. My findings indicated that the most accurate learners were those who checked at least some of their answers using strategies other than their first-choice strategies. For example, those learners might have applied their knowledge of divisibility rules or inverses to check answers that they had initially calculated another way. Nevertheless, most of the interviewees suggested that they simply checked each of their calculations by repeating their initial choice of strategy which was usually, but not always, a formal algorithm. We might infer that encouraging learners to develop a range of calculation skills, rather than focusing solely on formal algorithms, might increase their calculation accuracy.

The interviews revealed that most learners checked their answers as they went along but others delayed their checking procedures until they had finished their workbook. However, my findings relating to the optimum time for checking their calculations were inconclusive.

The literature also highlighted the importance of reflecting on the reasonableness of an answer (including Gojak, 2012; Keiser, 2012) but again there was a scarcity of recent studies addressing this topic. My findings provided only limited evidence that some Y6s did reflect on the reasonableness of their answers, but perhaps without realising it. More specifically, two of the less confident boys corrected answers without explicitly appearing to check their work or estimate their answers. In both cases, they appeared to draw on the lower level of number sense described by Berch (2005). It was perhaps interesting to note that there were no similar comments from girls or higher attainers, although a wider sample would need to be studied in the future to ascertain any possible trends in these findings. Again, encouraging learners to reflect on their answers might help them to spot errors and increase their calculation accuracy.

The interviews also raised questions regarding the role of out-of-school tutors, although the evidence was very limited. Just two of the interviewees revealed that they had mathematics tutors, but the actual figure may have been higher; around 10% of 11- to 16-year-olds had out of school tutors and mathematics was the most popular subject (Kirby, 2016, p. 2). At the time of writing, there was no governing body overseeing out-of-hours tuition in the UK which contrasted sharply with other, more heavily regulated businesses involving out of school care for children, such as child minders.

RQ3: Which are the most accurate calculation strategies for Year 6 learners completing age-related, context-free multi-digit written calculations?

Although the DfE’s (2013a) reforms prioritised the teaching of formal algorithms in schools, and this study showed that it was the most frequently selected strategy for each of the 16 calculations, it was not necessarily the most accurate or efficient strategy for those calculations. Indeed, although formal algorithms were significantly more accurate for just two of the 16 calculations (456 + 372 and 693 ÷

3), other strategies were significantly more accurate than formal algorithms for four of those 16 calculations (702 – 695, 27 x 63, 401 ÷ 25 and 517 ÷ 19).

From these findings we might infer that calculation accuracy could be improved for some calculations by employing strategies other than the formal algorithm. More specifically, counting-up was significantly more accurate than the subtraction formal algorithm for calculating 702 – 695, the grid method was significantly more accurate than long multiplication for calculating 27 x 63 and chunking down was significantly more accurate than long division for calculating 401 ÷ 25 and 517 ÷19. These findings were in line with the earlier findings reported by Borthwick and Harcourt-Heath (2007) who analysed the strategies chosen by learners for four calculations (546 + 423, 317 – 180, 56 x 24 and 222 ÷ 3); they reported that the learners were more accurate using counting-up or using a number line for calculating 317 – 180, they were more accurate using the grid method than long multiplication for calculating 56 x 24 and they were more accurate chunking up or chunking down than using long division for the calculation 222 ÷ 3. However, Borthwick and Harcourt-Heath did not state the significance of their findings. Since their 2007 study, age-related expectations had been raised (DfE, 2013a) and my research design took those changes into account. Nevertheless, from both studies we might infer the benefits of using strategies other than formal algorithms for at least some calculations.

RQ4: To what extent do calculation efficiency and understanding vary by gender, prior attainment and confidence?

Efficiency.

The perceived efficiency of formal algorithms (DfE, 2013b) was one of the key drivers behind the DfE’s (2013a) reforms. In this study, most of the interviewees justified at least one of their calculation choices by referring to its perceived efficiency, irrespective of whether they chose a formal algorithm. However, there was a marked gender difference in their responses. The Y6 girls tended to refer to the perceived efficiency of formal algorithms whereas the Y6 boys were more likely to refer to the perceived efficiencies of other strategies. These findings were in line with previous studies which had reported gender differences in calculation choices for other age groups (including Bailey et al., 2012; Carr & Davis, 2001; Fennema et al., 1998). Again, it was worth reflecting on my findings considering Villalobos’ (2009) theoretical model; if the Y6 girls associated too closely their early mathematical success with the perceived efficiency of formal algorithms then they might struggle later in their schooling when faced with non-routine problems.

Moving on to consider the interview data relating to prior attainment and confidence, although twice as many of the comments from the higher attaining learners justified their choice of formal algorithms over other strategies based on their perceived efficiency, the figure for the low confidence learners was even higher; three times as many of their comments referred to the perceived efficiency of their chosen strategies. These findings implied an inverse relationship between calculation accuracy and perceived efficiency for learners using formal algorithms; the low confidence learners tended to be less accurate than the other learners, yet they were more likely to justify their selection of the formal algorithm based on its perceived efficiency. These findings raised concerns about reforms prioritising formal algorithms over and above other strategies based on their perceived efficiency (DfE, 2013b) whilst actively discouraging Y6 learners from making efficient use of calculators in their classroom (DfE, 2013a.). Choosing a strategy based on its perceived efficiency may have reduced the chances of reaching the correct answer for some of the learners in this study.

***Understanding*.**

The literature review highlighted the importance of understanding for developing both number sense and calculation fluency (including Gilmore & Bryant, 2008; McIntosh et al., 1992; Russell, 2000) as well as the heated debates regarding the relative importance of conceptual understanding and procedural fluency for young mathematicians (Klein, 2007). The DfE's (2013a) reforms, reflecting more recent attempts to reconcile both viewpoints, were built around developing conceptual understanding and procedural fluency up to Y6. However, the literature review also noted the scarcity of suitable measuring instruments for understanding (Jones et al, 2013). Hence the decision was made to structure my analysis of the interviews around the four aspects of understanding suggested by Kilpatrick et al., (2001). The interviews indicated that the majority of the Y6 learners did not fully understand their calculation strategies. There were gender differences too. More of the boys than the girls demonstrated their conceptual understanding of subtraction, multiplication and division. The differences were most notable for multiplication and division, almost twice as many boys demonstrated their conceptual understanding than the girls.

All the higher attaining learners satisfied the criteria for demonstrating their conceptual understanding of addition and subtraction, but there was a marked drop for multiplication and division. Moreover, some of their comments also raised a possible lack of knowledge regarding the formal algorithm for division; several learners reported that it was only suitable for multi-digit dividends and single-digit divisors. Although their comments indicated the benefits of including a range of division calculations in the study, by enabling such misconceptions to be identified, I suggest that they also raised concerns regarding the ability of Y6 learners to manage the curriculum expectation of performing division calculations with multi-digit divisors (DfE, 2013a).

Most of the low confidence learners were also able to satisfy the criteria for demonstrating their conceptual understanding of addition and subtraction, but their numbers were even lower for multiplication and division than the higher attaining learners. Most of the low confidence interviewees chose the grid method for multiplication but many struggled to calculate the partial products required by that process. Moreover, some of the low confidence learners who favoured the grid method were unfamiliar with applying it to three-digit numbers. This finding again raised concerns regarding the multi-digit calculation demands introduced under the reforms (DfE, 2013a). Furthermore, by adopting a research design which accommodated a range of calculations across the four operations, these findings highlighted the benefits of such an approach for exposing misunderstandings in mathematics.

Reflection on the MMR Research Design

Focusing on the first cohort of Y6 learners studying under the reformed curriculum, this study adopted a sequential explanatory MMR design which explored the extent that gender, prior attainment and confidence predicted calculation fluency. The literature review revealed that the existing calculation studies tended to adopt either small-scale choice/no choice research designs, which restricted the possible range of strategies under consideration, or large-scale surveys which restricted the opportunities for exploring the reasons behind their findings. In both cases, the knowledge base often relied on studies addressing just one or two of the four calculation operations, making comparisons of calculation fluency between operations almost impossible. By adopting a sequential explanatory MMR design, I was able to conduct a large-scale survey in the first phase across the four operations. This approach enabled the consideration of similarities and differences between the strategies for the four operations, such as the finding that some learners had limited understanding about the suitability of some of their preferred strategies for certain multi-digit multiplication and division calculations. The findings from the first phase informed the selection of a smaller number of interviewees for the second phase. The interviews offered an opportunity to identify additional factors and explore the reasons behind the decision-making of the learners. They indicated that knowledge of alternative strategies and the marking scheme for SATs as well as checking procedures were also factors affecting calculation fluency.

Although the literature review reported an association between mathematical anxiety and calculation skills (including Krinzinger et al., 2009; Thomas & Dowker, 2016), there was a scarcity of data relating exploring any possible correlation between mathematical confidence and calculation. By including confidence as one of the three predictor variables in this study, my research design enabled light to be shed on the importance of confidence in predicting calculation accuracy and indicating the degree of willingness to deviate from choosing formal algorithms. My findings revealed that confidence was a significant predictor of both calculation accuracy and use of formal algorithms. Moreover, it was a more powerful predictor than either gender or prior attainment.

The literature review highlighted several existing calculation studies which either collected first-hand data using work sheets or workbooks for their participants (including Bailey et al., 2012; Carr & Davis, 2001), or studies using secondary data from SATs or national surveys (including Borthwick & Harcourt-Heath, 2007; Johnson, 1987). However, none of those data collection instruments was suitable for my study which followed recent curriculum reforms (DfE, 2013a). The review's findings informed the development of a new research instrument, for this study a calculation workbook featuring a variety of calculations for each of the four operations, which has subsequently been adopted by some of the schools in this study as part of their ongoing internal assessments.

Challenges arising during the study.

Data collection.

The workbooks fulfilled both of their roles. For the quantitative aspects, they provided reliable data regarding accuracy and use of formal algorithms. They also offered a useful starting point for discussions in the interviews because learners could use their own work as prompts during the discussions. Moreover, the methodology chapter set out the steps taken to address the validity and reliability of the workbook. However, in future studies it would be preferable for members of the research team to deliver the research instruments themselves rather than rely on classroom teachers. In this study I attempted to overcome my concerns raised about possible inconsistencies when class teachers administered the workbooks to their classes by scripting the process, but it was possible that not all of the teachers followed the processes precisely as scripted.

The research design placed enormous pressure on myself to ensure that the workbooks were returned promptly in readiness for the interviews. The logistics involved, especially arranging the delivery and collection of the materials across the 19 schools, took a considerable amount of my time. Again, a larger research team and administrative support would have reduced this level of pressure. Also, a research team would enable much more extensive checking procedures for the coding and uploading of data than was possible for a study led by a single researcher.

The workbook and confidence questionnaire, which were printed on both sides of a single sheet of A4, were economical to produce and efficient to process too. There were no obvious issues regarding their completion, the learners did not appear to miss our calculations in their workbooks and they seemed to respond well to the blend of smiling face images and the accompanying text on the Likert-style scale for the confidence questionnaires.

Limitations of the study.

One of the main limitations of this study was its cross-sectional nature. However, it was beyond the scope of this thesis to adopt a longitudinal approach, tracking the Y6 learners through their secondary schooling and beyond to monitor their subsequent mathematical development. My literature review discussed several studies which adopted a longitudinal approach (Bailey et al., 2012; Borthwick & Harcourt-Heath, 2007, 2010, 2011; Carr & Alexeev, 2011; Fennema et al., 1998). Their research designs enabled their researchers to consider trends over an extended period. In a future study, I would like to extend my studies to lead a longitudinal, large-scale project addressing the research question "Which are the most influential factors affecting the performance of high attaining Y6 girls in GCSE mathematics?" By adopting a longitudinal approach, which enabled the comparison of the characteristics of high attainers at Y6 and GCSE, I would be able to fully explore the predictions of Villalobos' (2009) model which suggested that high performing girls at primary level might struggle with their later schooling.

Implications of the findings.

Beginning with the implications for researchers, the literature review for this thesis highlighted that many previous studies tended to focus on either one or two operations (including Bailey et al., 2012; Carr & Davis, 2001; Fennema et al. 1998). In contrast, my study addressed each of the four operations. This research design led to findings which crossed the boundaries between the operations. For example, my findings highlighted misconceptions regarding the suitability of multiplication and division strategies for some multi-digit calculations. If future researchers continue to focus on single operations, they might overlook findings relevant to more than one operation.

Although previous studies reported an association between mathematical anxiety and calculation skills ((Krinzinger et al., 2009; Thomas & Dowker, 2000), my findings indicated an association between confidence and calculation fluency. However, even though the literature revealed a scarcity of confidence and calculation studies, and a very limited selection of possible confidence measures for my study (Eccles et al., 1993; Stankov et al., 2012; Vermeer et al. 2000; Walker, 2013), my findings indicated that confidence was a more significant factor regarding fluency than either gender or prior attainment. Although confidence may have been overlooked in previous studies, my findings suggested a strong case for its inclusion in future research.

The literature review revealed several reports highlighting the importance of checking answers (Cockcroft, 1982; Dowker, 1992; McIntosh et al. 1992; Russell, 2000), estimating (Gojak, 2012;

Keiser, 2010) and reflecting on the reasonableness of answers (Cockcroft, 1982; Threadgill-Sowder, 1984), but there was a scarcity of recent reports addressing the most effective ways to do so. My findings indicated that learners who checked their answers using different strategies than their first-choice approach tended to work more accurately than other learners. Hence there appeared to be a strong argument for future studies to further investigate checking procedures, estimating and reflecting on the reasonableness of answers in more depth, especially the most effective checking strategies and the optimum time for checking work.

Moving on to the implications for policymakers from my findings, if calculation fluency required accuracy, efficiency, flexibility and conceptual understanding, then far too many of the Y6 learners in this study failed to demonstrate it. Formal algorithms were the most frequently-selected strategy for each of the 16 calculations, even though they were not necessarily the most accurate strategies for those calculations. By prioritising formal algorithms at the expense of other approaches, many of the learners in this study were apparently working inefficiently and following procedures which led to avoidable errors. Moreover, the finding that the girls were even more likely than the boys to choose formal algorithms was particularly concerning considering the theoretical model proposed by Villalobos (2009) which raised concerns about those girls managing the curriculum shift towards a problem-solving curriculum as they progress through their schooling.

The findings also raised concerns about the age-related expectations across the four operations (DfE, 2013a). Many of the learners, especially the less confident learners, struggled to cope with the multiplication and division calculations in this study. Other learners were unsure whether their strategies were applicable to larger numbers. Ensuring that all Y6 learners have sufficient opportunities to rehearse their strategies requires considerable classroom time, and brings into question the relevance of learning such strategies when many learners can easily access a calculator. Rather than focusing on performing formal algorithms, schools could encourage learners to develop their metacognition skills. In particular, they could support their learners to identify whether the most appropriate approach for an individual calculation would involve working mentally, performing a pencil-and-paper approach or fetching a calculator. Such an approach should enable learners to avoid selecting inefficient written approaches when they could easily work out an answer in their heads and allow them more time to work with real-life problems involving large numbers if they used a calculator when necessary.

Personal Reflections on the Research Journey

As a researcher, I feel that I have developed a much stronger understanding regarding the data collection and analysis of both quantitative and qualitative data as well as appreciating the gains from working with both aspects of data collection within a single study. However, considerable time was spent developing my skills for a single research project which was attracting interest from schools. If I had worked within a larger team, then I might have been able to complete the data analysis and disseminate my results within a reduced timeframe. An alternative approach might have involved identifying key points during the study for submitting papers on particular aspects of the ongoing work; by taking more than one approach towards the dissemination aspect of the study, I would have been able to share my findings much more quickly. In other words, a study addressing calculation fluency has highlighted the importance of 'dissemination fluency' for me too.

As a teacher, my beliefs regarding calculation fluency have evolved during this study. At the start, I was curious about the ways that learners would respond to the forthcoming reforms in our classrooms. When I started to analyse my data and had the time to reflect on my findings and the literature that I had read, I began to realise the potentially far-reaching consequences of the curriculum reforms (DfE, 2013a). As educators, we have a responsibility to ensure that the learners we have been entrusted to teach receive the very best opportunities to learn and develop as future citizens; this thesis has highlighted for me the importance of thoroughly researching ideas and policies prior to their national rollout. Working on this thesis has strengthened my interests in curriculum development, collecting evidence and influencing policymakers. Conducting a PhD study offered a demanding academic exercise. Maximising the potential of my findings is my next challenge.

References

Advisory Committee on Mathematics Education (ACME). (2012). *Raising the bar: Developing able young mathematicians.* Retrieved from http://acme-uk.org/media/10498/raisingthebar.pdf

Aldridge, A., & Levine, K. (2001). Surveying the social world. *Order*, *12*, 196.

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, *12*(5), 385-390.

Anghileri, J. (2000). *Teaching number sense*. London: A&C Black.

Anghileri, J. (2001). A study of progression in written strategies for division. *Support for Learning*, *16*(1), 17-22.

Anghileri, J. (2006). A study of the impact of reform on students' written calculation methods after five years' implementation of the National Numeracy Strategy in England. *Oxford Review of Education, 32*(3), 363-380.

Anghileri, J., Beishuizen, M., van Putten, K. and Snijders, P. (1999, September). *A comparison of English and Dutch Year 5 pupils' calculation strategies for division.* Paper presented at British Educational Research Association Annual Conference, University of Sussex, Brighton.

Ansari, D. (2014). Understanding Developmental Dyscalculia: A Math Learning Disability [webinar]. Retrieved from https://www.ldatschool.ca/wp-content/uploads/2015/03/Transcript_Understanding-Developmental-Dyscalculia-Webinar.pdf

Back, J. (2014, February). *Early number sense*. Retrieved from https://nrich.maths.org/10737

Bailey, D., Littlefield, A., & Geary, D. (2012). The codevelopment of skill at and preference for use of retrieval-based processes for solving addition problems: Individual and sex differences from first to sixth grades. *Journal of Experimental Child Psychology, 113*, 78-92.

Baroody, A., Torbeyns, J., & Verschaffel, L. (2009). Young children's understanding and application of subtraction-related principles. *Mathematical Thinking and Learning*, *11*(1-2), 2-9.

Bazeley, P. (2002). Computerized data analysis for mixed methods research. In A. Tashakorri, & C. Teddlie (Eds.), *Handbook of Mixed Methods for the Social and Behavioral sciences* (pp. 385-422). Thousand Oaks, CA: Sage.

Berch, D. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of learning disabilities*, *38*(4), 333-339.

Berch, D. B., & Mazzocco, M. M. (2007). Why is math so hard for some children? *The Nature and Origins of Mathematical Learning Difficulties and Disabilities.* Baltimore, MD: Paul H. Brookes Pub. Co.

Berriman, R., Hawksworth. J., Kelly, C., & Foyster, R. (2017, March). Will robots steal our jobs? The potential impact of automation on the UK and other major economies. *UK Economic Outlook*. Retrieved from https://www.pwc.co.uk/economic-services/ukeo/pwc-uk-economic-outlook-full-report-march-2017-v2.pdf

Bisson, M., Gilmore, C., Inglis, M., & Jones, I. (2016). Measuring conceptual understanding using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education*, *2*(2), 141-164.

Books.google.com. (2017). *Google Ngram Viewer*. Retrieved from https://books.google.com/ngrams

Borthwick, A., & Harcourt-Heath, M. (2007). Calculation strategies used by Year 5 children. *Proceedings of the British Society for Research into Learning Mathematics*, *27*(1), 12-17.

Borthwick, A., & Harcourt-Heath, M. (2010). Calculating: What can Year 5 children do? *Proceedings of the British Society for Research into Learning Mathematics*, *30*(3), 13-18.

Borthwick, A., & Harcourt-Heath, M. (2012). Calculating: What can Year 5 children do now? *Proceedings of the British Society for Research into Learning Mathematics*, *32*(3), 25-30.

Borthwick, A., & Harcourt-Heath, M. (2015). Calculating: How have Year 5 children's strategies changed over time? *Proceedings of the British Society for Research into Learning Mathematics*, *34*(3), 1-6.

Borthwick, A., & Harcourt-Heath, M. (2016). Calculation strategies for Year 5 children: Ten years on. *Proceedings of the British Society for Research into Learning Mathematics*, *36*(3), 1-6.

Borthwick, A., Harcourt-Heath, M., & Keating, R. (2014). Calculating: What can Year 8 children do? *Proceedings of the British Society for Research into Learning Mathematics, 34*(1), 25-30.

British Educational Research Association (BERA). (2011). *Ethical guidelines for educational research*. Retrieved from https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2011

Brown, M. (1987). Number. In K. Hart, M. Brown, D. Kerslake, D, Kűchemann, & G. Ruddock (Eds.), *Chelsea Diagnostic Mathematics Tests* (pp. 73-70). Windsor, England: NFER-Nelson.

Brown, M., Brown, P., & Bibby, T. (2008). "I would rather die": Attitudes of 16-year-olds towards their future participation in mathematics. *Research in Mathematics Education, 10*(1), 2–18.

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, *6*(1), 97-113.

Burgess, H., Sieminski, S., & Arthur, L. (2006). *Achieving your doctorate in education.* London, England: Sage.

Cai, J., & Hwang, S. (2002). Generalized and generative thinking in US and Chinese students' mathematical problem solving and problem posing. *The Journal of Mathematical Behavior*, *21*(4), 401-421.

Caplan, P., MacPherson, G.., & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? A multilevel critique with new data. *American Psychologist*, *40*(7), 786.

Carr, M., & Davis, H. (2001). Gender differences in arithmetic strategy use: A function of skill and preference. *Contemporary Educational Psychology*, *26*, 330-347.

Carr, M. & Alexeev, N. (2011). Fluency, accuracy and gender predict developmental trajectories of arithmetic strategies. *Journal of Educational Psychology*. *103*(3), 617-631.

Case, R. (1998, April). *A psychological model of number sense and its development*. Presented at annual meeting of the American Educational Research Association, San Diego, CA.

Che, M., Wiegert, E., & Threlkeld, K. (2012). Problem solving strategies of girls and boys in single-sex mathematics classrooms. *Educational Studies in Mathematics, 79,* 311-326.

Cockcroft, W. H. (1982). *Mathematics counts: Report of the committee of inquiry into the teaching of mathematics in schools*. London, England: Her Majesty's Stationery Office.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). London, England: Routledge.

Creswell, J. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). London, England: Pearson.

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

Crotty, M. (1998). The foundations of social research: Meaning and perspective in the research process. London, England: Sage.

Cuoco, A., Goldenberg, E. P., & Mark, J. (1996). Habits of mind: An organizing principle for mathematics curricula. *The Journal of Mathematical Behavior, 15*(4), 375-402.

Denscombe, M. (2010). *The good research guide for small-scale social research projects* (4th ed.). Maidenhead, England: Open University Press.

Department for Children, Schools and Families (DCSF). (2009). *Assessing pupils' progress (APP) in mathematics: Criteria levels 1 to 8.* Nottingham, England: Department for Children, Schools and Families.

Department for Education (DfE). (2013a). *National curriculum in England: Framework document.* Retrieved from https://www.gov.uk/government/publications/national-curriculum-in-england-primary-curriculum

Department for Education (DfE). (2013b). *Education minister Elizabeth Truss speaks at the North of England Education Conference.* Retrieved from https://www.gov.uk/government/speeches/education-minister-elizabeth-truss-speaks-at-the-north-of-england-education-conference

Department for Education (DfE). (2013c). National curriculum assessments at Key Stage 2 in England. Retrieved from https://docs.google.com/spreadsheets/d/1vlmVznYbdn0vB47OHOFIm0BhZaDbQW_vINfe1xqZr5g/edit#gid=0

Department for Education (DfE). (2013d). *National curriculum assessments and review outcomes at key stage 2, including measures of progress between key stages 1 and 2*. Retrieved from https://www.gov.uk/government/collections/statistics-key-stage-2

Department for Education (DfE) (2014a). *Schools, pupils and their characteristics: January 2014*. Retrieved from: https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2014

Department for Education (DfE). (2017). *2017 national curriculum tests key stage 2 mathematics paper 1: Arithmetic*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/614656/STA177736e_2017_key_stage_2_mathematics_paper_1_arithmetic.pdf

Department for Education and Employment (DfEE). (1999) *Framework for teaching mathematics from Reception to Year 6.* Nottingham, England: Department for Education and Employment.

Department for Education and Skills. (DfES). (2007) *Guidance paper: Calculation.* Retrieved from http://webarchive.nationalarchives.gov.uk/20110206001854/https://nationalstrategies.standards.dcsf.gov.uk/node/47364?uc=force_uj

Dowker, A. (1992). Computational strategies of professional mathematicians. *Journal for Research in Mathematics Education, 23*(1), 45-55.

Dowker, A. (1998). Individual differences in normal arithmetical development. In C. Donlan (Ed.), *Studies in developmental psychology. The development of mathematical skills* (pp. 275-302). Hove, England: Psychology Press/Taylor & Francis.

Dowker, A. (2004). *What works for children with mathematical difficulties?* (Vol. 554). Nottingham, England: DfES Publications.

Dowker, A. (2014). Young children's use of derived fact strategies for addition and subtraction. *Frontiers in human neuroscience*, *7,* 924.

Dowker, A., Sarkar, A., & Looi, C. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in psychology*, *7,* 508.

Eccles, J., Wigfield, A., Harold, R. & Blumenfeld, P. (1993). Age and gender differences in children’s self and task perceptions during elementary school. *Child Development, 64*(3), 830-847.

Else-Quest, N., Hyde, J., & Linn, M. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*(1), 103-127.

Fagginger Auer, M. F., Hickendorff, M., Van Putten, C. M., Béguin, A. A., & Heiser, W. J. (2016). Multilevel latent class analysis for large-scale educational assessment data: Exploring the relation between the curriculum and students’ mathematical strategies. *Applied Measurement in Education*, *29*(2), 144-159.

Fennema, E. (2000, May). *Gender and mathematics: What is known and what I wish was known.* Paper presented at the Fifth Annual Forum of the National Institute for Science Education, Michigan, US.

Fennema, E., Carpenter, T. P., Jacobs, V. R., Franke, M. L., & Levi, L.W. (1998). A longitudinal study of gender differences in young children's mathematical thinking. *Educational Researcher*, *27*(5), 6-11.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London, England: Sage.

Fielker, D. (2007). Addition and subtraction, and algorithms in general. *Mathematics Teaching, 204*, 3-5.

Fink, A., & Kosecoff, J. (1998). *How to conduct surveys: A step-by step guide*. London, England: Sage.

Fishbein, M. M., & Azjen, I. F. (1975). *Belief, intention, attitude, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Fowler, F. (2002). *Survey Research Methods* (3rd ed.). London, England: Sage.

Gallagher, A., De Lisi, R., Holst, P.., McGillicuddy-De Lisi, A., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology*, *75*(3), 165-190.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: a five-year prospective study. *Journal of Educational Psychology*, *104*(1), 206.

Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology*, *15*(4), 380.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education*, *33*(1), 18-28.

Gibson, D. (2010). Quantitative research as a method of inquiry in education. In D. Hartas, (Ed.), *Educational research and inquiry: Qualitative and quantitative approaches* (pp. 65-81). London, England: A&C Black.

Gill, T. (2015). *GCSE uptake and results by gender 2004-2013.* Cambridge, England: Cambridge Assessment.

Gillham, B. (2007). *Developing a questionnaire* (2nd ed.). London, England: Continuum.

Gilmore, C., & Bryant, P. (2008). Can children construct inverse relations in arithmetic? Evidence for individual differences in the development of conceptual understanding and computational skill. *British Journal of Developmental Psychology, 26*(3), 301-316.

Gojak, L. (2012, November). *Fluency: Simply fast and accurate? I think not!* Retrieved from https://www.nctm.org/News-and-Calendar/Messages-from-the-President/Archive/Linda-M_-Gojak/Fluency_-Simply-Fast-and-Accurate_-I-Think-Not!/

Gojak, L. (2013). *The power of a good mistake*. Retrieved from https://www.nctm.org/News-and-Calendar/Messages-from-the-President/Archive/Linda-M_-Gojak/The-Power-of-a-Good-Mistake/

Gorard, S. A. C. (2001). *Quantitative methods in educational research: The role of numbers made easy.* London, England: Continuum.

Gorard, S. (2007). The dubious benefits of multi-level modeling. *International Journal of Research & Method in Education*, *30*(2), 221-236.

Gorard, S. A. C. (2015). Rethinking ‘quantitative’methods and the development of new researchers. *Review of Education.*, *3*(1), 72-96.

Greene, J., Caracelli, V., & Graham, W. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*(3), 255-274.

Gross, J. (2007). Supporting children with gaps in their mathematical understanding: The impact of the National Numeracy Strategy (NNS) on children who find mathematics difficult. *Educational and Child Psychology, 24*(2), 146–156.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of Qualitative Research*, *2,* 163-194.

Hartas, D. (Ed.). (2010). *Educational research and inquiry: Qualitative and quantitative approaches*. London, England: A&C Black.

Heilbronner, N. N. (2013). The STEM pathway for women: What has changed? *Gifted Child Quarterly*, *57*(1), 39–55.

Hickendorff, M., van Putten, C., Verhelst, & Heiser, W. (2010). Individual differences in strategy use on division problems: Mental versus written computation. *Journal of Educational Psychology, 102*(2), 438-452.

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. *Conceptual and Procedural Knowledge: The Case of Mathematics, 2*, 1-27.

Hodgen, J., Adkins, M., Ainsworth, S., & Evans, S. (2019). *Catch Up® Numeracy: Evaluation report and executive summary*. Retrieved from https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Catch_Up_Numeracy.pdf

Howe, K. R. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational Researcher, 17*(8), 10-16.

Howitt, D., & Cramer, D. (2005). *Introduction to SPSS in psychology: With supplements for releases 10, 11, 12 and 13.* London, England: Pearson Education.

Ivankova, N., Creswell, J., & Stick, S. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, *18*(1), 3-20.

Johnson, R., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14-26.

Johnson, S. (1987). Early-developed sex differences in science and mathematics in the United Kingdom. *The Journal of Early Adolescence*, *7*(1), 21-33.

Jones, I., Inglis, M., Gilmore, C., & Hodgen, J. (2013, July 28th – August 2nd). Measuring conceptual understanding: The case of fractions. In Lindmeier, A. & Heinze, A. (Eds.). Paper presented at the Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education, (3), 113-120. Kiel, Germany: PME.

Keiser, J. M. (2010). Shifting our computational focus. *Mathematics Teaching in the Middle School, 16*(4), 217-223.

Keiser, J. M. (2012). Students' strategies can take us off guard. *Mathematics Teaching in the Middle School*, *17*(7), 418-425.

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

Kirby, P. (2016). *Shadow Schooling: private tuition and social mobility in the UK*. London, England: The Sutton Trust.

Klein, D. (2007). A quarter century of US ‘math wars’ and political partisanship. *British Society for the History of Mathematics Bulletin*, *22*(1), 22-33.

Krinzinger, H., Kaufmann, L., Dowker, A., Thomas, G., Graf, M., Nuerk, H., & Willmes, K. (2007). German version of the math anxiety questionnaire (FRA) for 6-to 9-year-old children. *Zeitschrift fur Kinder-und Jugendpsychiatrie und Psychotherapie*, *35*(5), 341-351.

Krinzinger, H., Kaufmann, L., & Willmes, K. (2009). Math anxiety and math ability in early primary school years. *Journal of Psychoeducational Assessment*, *27*(3), 206-225.

Kuhn, T. S. (2012). *The structure of scientific revolutions*. Retrieved from http://projektintegracija.pravo.hr/_download/repository/Kuhn_Structure_of_Scientific_Revolutions.pdf

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, *124*(1), 83-97.

Lindberg, S., Hyde, J., Petersen, J., & Linn, M. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, *136*(6), 1123-1135.

Loosveldt, G. (2008). Face to face interviews. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 201-220). London: Erlbaum Associates, Inc.

Ma, L. (1999*). Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Maxwell, J. (1997). Designing a qualitative study. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 69-100). Thousand Oaks, CA: Sage.

McClure, L. (2014, April). *Developing number fluency – what, why and how*. Retrieved from https://nrich.maths.org/10624

McIntosh, A., Reys, B. J., & Reys, R. E. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics*, *12*(3), 2-44.

McWhinnie, S., & Fox, C. (2013). *Advancing women in mathematics: Good practice in university departments.* London, England: London Mathematical Society.

Middleton, J., & Spanias, P. (1999). Motivation for achievement in mathematics: Findings, generalizations, and criticisms of the research. *Journal for Research in Mathematics Education*, *30*, 65-88.

Morgan, D. (2007). Paradigms lost and pragmatism regained methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, *1*(1), 48-76.

Morse, J. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, *40*, 120–123.

Mullis, I. V. S., Martin, M. O., Foy, P. (2008). TIMSS 2007 International mathematics report: Findings from IEA’s Trends in International Mathematics and Science Study at the fourth and eighth grade. *International Association for the Evaluation of Educational Achievement*. Retrieved from https://timss.bc.edu/TIMSS2007/PDF/TIMSS2007_InternationalMathematicsReport.pdf

Neidorf, T., Binkley, M., Gattis, K., & Nohara, D. (2006). Comparing mathematics content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments. Technical report. NCES 2006-029. *National Center for Education Statistics*. Retrieved from http://nces.ed.gov/pubs2006/2006029.pdf

Neuville, E., & Croizet, J. (2007). Can salience of gender identity impair math performance among 7–8 years old girls? The moderating role of task difficulty. European Journal of Psychology of Education, 22*(*3), 307-316.

Nunes, T., Bryant, P., Sylva, K., & Barros, R. (2009). *Development of maths capabilities and confidence in primary school.* Research Report DCSF RR118. London, England: Department for Children, Schools and Families.

Office for Standards in Education, Children’s Services and Skills (Ofsted). (2011). *Good practice in primary mathematics: Evidence from 20 successful schools*. Manchester, England: Ofsted.

Onwuegbuzie, A., & Johnson, R. (2006). The validity issue in mixed research. *Research in the Schools, 13*(1), 48-63.

Oppenheim, A. (1992). *Questionnaire design and aptitude measurement*. London, England: Heinemann.

Organisation for Economic Co-operation and Development (OECD). (2012). *PISA in Focus 14: What kinds of careers do boys and girls expect for themselves?* Retrieved from http://www.oecd.org/pisa/pisaproducts/pisainfocus/49829595.pdf

Organisation for Economic Co-operation and Development (OECD). (2013). *Ready to learn: Students' engagement, drive and self-beliefs – volume III.* Retrieved from http://www.oecd.org/pisa/keyfindings/PISA2012-Vol3-Chap4.pdf

Organisation for Economic Co-operation and Development (OECD). (2014). *PISA 2012 results in focus. What 15 year olds know and what they can do with what they know*. Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

Oxford University Press (OUP). (2013). *Oxford English Dictionary Online*. Retrieved from http://www.oed.com/

Pampaka, M., Kleanthous, I., Hutcheson, G. D., & Wake, G. (2011). Measuring mathematics self-efficacy as a learning outcome. *Research in Mathematics Education, 13*(2), 169-190.

Peterson, P., & Fennema, E. (1985). Effective teaching, student engagement in classroom activities, and sex-related differences in learning mathematics. *American Educational Research Journal*, *22(*3), 309-335.

Plunkett, S. (1979). Decomposition and all that rot. *Mathematics in School*. *8(3*), 2-7.

Pring, R. (2000). *Philosophy of educational research*. London, England: Continuum.

R Core Team. (2016). R: A language and environment for statistical computing. *R Foundation for Statistical Computing.* Vienna, Austria. Retrieved from https//www.R-project.org/

Richardson, F., & Suinn, R. (1972). The mathematics anxiety rating scale: psychometric data. *Journal of Counselling Psychology*, *19*(6), 551.

Richie, J., Spencer, L. (1994). Qualitative data analysis for applied policy research. In A. Bryman & R. Burgess (Eds.), *Analysis of qualitative data* (pp. 187-208). London, England: Routledge.

Robson, C. (2002). *Real World Research* (2nd ed.). Oxford, England: Blackwell.

Rubio, D., Berg-Weger, M., Tebb, S.., Lee, E.., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, *27*(2), 94-104.

Russell, S. J. (2000). *Developing computational fluency with whole numbers in the elementary grades.* Retrieved from https://investigations.terc.edu/inv2/wp-content/uploads/2017/10/Developing-Computational-Fluency-with-hole-Numbers-in-the-Elementary-Grades.pdf

Ruthven, K. (1995). Beyond common sense: Reconceptualizing national curriculum assessment. *The Curriculum Journal, 6*(1), 5-28.

Ruthven, K. (2009). Towards a calculator-aware mathematics curriculum. *Mediterranean Journal for Research in Mathematics Education, 8*(1), 1-14.

Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, *126*(1), 71.

Singer, E. (2008). Ethical issues in surveys. In E. de Leeuw., J. Hox., D. Dillan (Eds.), *International Handbook of Survey Methodology* (pp. 78-96). London, England: Lawrence Erlbaum Associates.

Srivastava, A., & Thomson, S. B. (2009). Framework analysis: A qualitative methodology for applied policy research. *Journal of Administration and Governance*, *4*(2), 72-79.

Standards and Testing Agency (STA). (2016a). *2016 National curriculum assessments Key Stage 2 mathematics test mark schemes Paper 1: Arithmetic, Paper 2: Reasoning, Paper 3: Reasoning.* Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524746/2016_ks2_mathematics_markschemes_PDFA.pdf

Standards and Testing Agency (STA). (2016b). *2016 National curriculum tests Key Stage 2 mathematics administering the mathematics test Paper 1: arithmetic.* Retrieved from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524057/2016_Key_stage_2_mathematics_test_Paper_1_admin_instructions_PDFA.pdf

Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences*, *22*(6), 747-758.

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, 404-411.

Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, *18*(6), 565-579.

Star, J. R., Rittle-Johnson, B., & Durkin, K. (2016). Comparison and explanation of multiple strategies: One example of a small step forward for improving mathematics education. *Policy Insights from the Behavioral and Brain Sciences*, *3*(2), 151-159.

Taber, K. (2007). *Classroom-based research and evidence-based research: A guide for teachers.* London, England: Sage.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. London, England: Sage.

Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & Teddlie, C (Eds), *Handbook of Mixed Methods in Social &Behavioral Research* (pp. 3-50). Thousand Oaks, CA: Sage

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research*, *1*(1), 77-100.

Thomas, G., & Dowker, A. (2000, September). Mathematics anxiety and related factors in young children. In *British Psychological Society Developmental Section Conference*.

Thompson, I. (2008). Deconstructing calculation methods, part 4: Division. *Mathematics Teaching Incorporating Micromath*, *208*, 6-8.

Threadgill-Sowder, J. (1984). Computational estimation procedures of school children. *The Journal of Educational Research*, *77*(6), 332-336.

Tomlinson, P. (1989). Having it both ways: Hierarchical focusing as research interview method. *British Educational Research Journal*, *15*(2), 155-176.

Torbeyns, J., & Verschaffel, L. (2013). Efficient and flexible strategy use on multi-digit sums: A choice/no-choice study. *Research in Mathematics Education*, *15*(2), 129-140.

Vermeer, H. J., Boekaerts, M., & Seegers, G. (2000). Motivational and gender differences: Sixth-grade students' mathematical problem-solving behavior. *Journal of Educational Psychology*, *92*(2), 308.

Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. K. Lester, Jr. (Ed.), Second handbook of research on mathematics teaching and learning (pp. 557–628). New York: Information Age Publishing.

Villalobos, A. (2009). The importance of breaking set: Socialized cognitive strategies and the gender discrepancy in mathematics. *Theory and Research in Education*, *7*(1), 27-45.

Walker, M., Straw, S., Jeffes, J., Sainsbury, M., Clarke, C., & Thom, G. (2013). *Evaluation of the Mathematics Specialist Teacher (MaST) programme*. Retrieved from https://www.nfer.ac.uk/media/2085/pmsz01.pdf

Williams, P. (2008). *Independent review of mathematics teaching in early years settings and primary schools: Final report - Sir Peter Williams 2008.* Nottingham, England: Department for Children, Schools and Families.

Wolfram, C. (2014, February 23). The UK needs a revolution in the way maths is taught. *The Guardian.* Retrieved from https://www.theguardian.com/education/2014/feb/23/maths-teaching-revolution-needed-conrad-wolfram

Yin, R. (2009) *Case study research: Design and methods* (4th ed.). London, England: Sage.

Zohar, A., & Gershikov, A. (2008). Gender and performance in mathematical tasks: Does the context make a difference? *International Journal of Science and Mathematics Education*, *6*(4), 677-693.

Appendix A

Marking guidance prioritising use of the formal algorithm (STA, 2016a, p. 13)

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524746/2016_ks2_mathematics_markschemes_PDFA.pdf

Qu.	Requirement	Mark	Additional guidance
23	Award **TWO** marks for the correct answer of 3,266 If the answer is incorrect, award **ONE** mark for the formal method of long multiplication with no more than **ONE** arithmetical error, e.g. • 71 × 46 426 2840 3260 *(error)* **OR** • 71 × 46 426 2440 *(error)* 2866	**Up to 2m**	Working must be carried through to reach a final answer for the award of **ONE** mark. **Do not** award any marks if the error is in the place value, e.g. the omission of the zero when multiplying by tens: 71 × 46 426 284 *(place value error)* 710

Appendix B

Level Four and Level 5 Descriptors for 'Number and Calculation'

Level 5 APP descriptors: Number and calculation • Use known facts, place value, knowledge of operations and brackets to calculate including using all four operations with decimals to two decimal places • Use a calculator where appropriate to calculate fractions/percentages of quantities/measurements • Understand and use an appropriate non-calculator method for solving problems that involve multiplying and dividing any three digit number by any two digit number • Solve simple problems involving ordering, adding, subtracting negative numbers in context • Solve simple problems involving ratio and direct proportion • Apply inverse operations and approximate to check answers to problems are of the correct magnitude
Level 4 APP descriptors: number and calculation • Use a range of mental methods of computation with all operations • Recall multiplication facts up to 10 x 10 and quickly derive corresponding division facts • Use efficient written methods of addition and subtraction and of short multiplication and division • Multiply a simple decimal by a single digit • Solve problems with or without a calculator • Check the reasonable ness of results with reference to the context or size of the numbers

Note. Highlighted statements refer to 'problems'

Appendix C

Worked examples of the calculation strategies included in the final coding scheme

Chunk down

	24	
20	480	
-	200	(10x20)
	280	
-	200	(10x20)
	80	
-	80	(4x20)
	0	

Count back

↙	-5	↖	↙	-2	↖
695		700			702

702 - 695 = 7

517 - 19 - … - 19

517 ÷ 19 = 27r4

Count-up

↗	+300	↘	↗	+70	↘	↗	+2	↘
456		756			836			838

456 + 372 = 838

↗	+5	↘	↗	+2	↘
695		702			702

702 - 695 = 7

27 + … + 27

27 x 63 = 1701

↗	4x16	↘	↗	+1	↘
0		400		401	401

401 ÷25 = 16r1

Doubling

245 + 245 = 490
490 + 11 = 501
So 245 + 256 = 501

Expanded

	400	+	50	+	6
+	300	+	70	+	2
	700	+	120	+	8

So 456 + 372 = 828

	600	+	30	+	2
-	100	+	50	+	4
	400	+	70	+	8

So 632 - 154 = 478

	27
x	63
	21
	420
	60
	1200
	1701

Formal algorithm

	456
+	372
	$_{1}828$

	$3^{1}8^{1}2$
-	$^{2}199$
	183

Or

$^{2}3^{17}8^{1}2$
199
183

	27
x	63
	81
	1620
	1701

$$3\overline{)693}$$ = 231

Grid

	20	7
60	1200	420
3	60	21

	1200
	420
	60
+	21
	1701

Multiples of 10

2 x 46 = 92
So
20 x 46 = 920

48 ÷ 2 = 24
So
480 ÷ 20 = 240

Number bonds

76 + 24 = 100
So 500 - 76 = 424

Partition

400	+	300	=	700		300	-	100	=	200		
50	+	70	=	120		80	-	90	=	-10	*	
6	+	2	=	8		2	-	9	=	-7	*	
So 456	+	372	=	828		So 382	-	199	=	183		

*Or borrow ie100 to make 180-90=90 and 12-9=3

	27	x	63		1200
↙	↓	↓	↘		420
20x60	7x60	20x3	7x3		60
↙	↓	↓	↘	+	21
1200	420	60	21		1702

	600	90	3
÷3	↓	↓	↓
	200	30	1

So 693 ÷ 3 = 231

Round

300	+	532	-	1	=	832	-	1
					=	831		
So	299	+	532	=	831			

382	-	200	+	1	=	182	+	1
					=	183		
So	382	-	199	=	183			

35	x	100	=	3500
3500	-	35	=	3475
So	35	x	99	= 3475

400	÷	25	=	16
So	401	÷	25	= 16r1

Appendix D

Guidance for class teacher

Thank you for your support with this research project. The following notes highlight some key points which arose during the trialling of the workbooks:

1. Please allow 15 minutes for the confidence questionnaire and 40 minutes for pupils to complete their workbook. In the trial pupils could complete both tasks within a lesson.
2. Please ensure that the usual test procedures are followed, including covering up calculation displays and reminding pupils when they are halfway through their 40 minutes time allocation for the workbook as well as when they have 5 minutes remaining
3. The questionnaire was written for primary-aged pupils and was designed to be read aloud to them, one question at a time.
4. The calculations were designed for pupils working at Y6 age-related expectations. If any pupil falls significantly below those expectations, please use your professional judgement to decide whether or not they should participate. If a pupil usually has an adult helper, then that helper can support the pupil within the usual test guidelines

Thank you

Ems Lord

Appendix E

Expert validation form

Your name:

Thank you for agreeing to participate in the expert review of the pupil workbook items that I am developing for my PhD study. Your feedback is entirely confidential. Below is a description of my study, how I have defined key terms and a list of questions about the items in the pupil workbook. Please begin by familiarising yourself with this background information and its definitions, and then complete the reviewer questions in the subsequent section.

<u>My research project</u>

Following the introduction of our new mathematics curriculum (DfE, 2013), and its focus on teaching standard written calculation algorithms, I would like to investigate fluency in written calculations. In particular, I am interested in comparing the strategies chosen for a range of written calculations by gender and confidence levels.. I hope that my research will inform my work as a lecturer in mathematics education.

I have four research questions:

1: To what extent do gender, prior attainment or confidence predict use of the formal algorithm?

2: To what extent do gender, prior attainment and confidence predict calculation accuracy?

3: Which are the most accurate calculation strategies for year 6 pupils completing age-related, context-free multi-digit written calculations?

4: To what extent do calculation efficiency and understanding vary by gender, prior attainment and confidence?

To address the first three questions, I will be surveying the responses of a large number of Y6 pupils using a workbook, similar in style to a SATs paper. It contains 16 calculations; four per operation. The completed calculations will be checked for accuracy and choice of calculation strategies. Then, a smaller sample of pupils will be interviewed about their calculation choices.

<u>A context-free multi-digit written calculation</u>

For this study, a written calculation has been defined as any calculation where Y6 pupils working at age-related expectations would be reasonably expected to record some jottings, informal procedures or algorithms in their workbook. Calculators would be not allowed.

The calculations are context-free to avoid any unintentional gender stereotyping. For example, using a word problem referring to football or cooking could unintentionally trigger a gender response which previous researchers have shown can influence calculation responses.

Thank you for your help

1. Clarity of the workbook

This section refers to how understandable you feel the layout of the workbook is for pupils. Please rate it using the scale below. *If you have any ideas for how to clarify the layout, please note your suggestions beneath the scale.*

Not at all understandable	Somewhat understandable	Quite understandable	Extremely understandable

Suggestions:

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

1. Relevance

In this section, I would like to know how relevant you feel each question is for my research design. Please rate the relevance of each question by ticking the appropriate box. *If you have any ideas for improving the relevance of the calculation please note your suggestions beneath the scale.*

Question One: 456 + 372

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

…….
…….

Question Two: 203 + 401

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

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

Question Three: 245 + 256

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

………

……………………………………………………………………………………………

Question Four: 299 + 534

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

………

……………………………………………………………………………………………

Question Five: 632 - 154

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

………

……………………………………………………………………………………………

Question Six: 382 - 19

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

………

…………………………………………………………………………………………..

Question Seven: 702 - 695

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

………

……………………………………………………………………………………………

Question Eight: 500 - 76

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

……

…………………………………………………………………………………………..

Question Nine: 27 x 63

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

……

……………………………………………………………………………………………

Question Ten: 35 x 99

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

……

……………………………………………………………………………………………

Question Eleven: 20 x 46

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

……

…………………………………………………………………………………………..

Question Twelve: 568 x 34

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

……

……………………………………………………………………………………………

Question Thirteen: 517 ÷ 19

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

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

Question Fourteen: 480 ÷ 20

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

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

Question Fifteen: 693 ÷ 3

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

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

Question Sixteen: 401 ÷25

Not at all relevant	Somewhat relevant	Quite relevant	Extremely relevant

Suggestions:

……
…………………………………………………………………………

2. <u>Finally, please think about the workbook as a whole for a moment.</u>

I hope that this workbook fairly represents the four operations for written calculations without overlooking any key types of calculation for each operation. Please indicate if you feel key calculations have been overlooked or not fairly represented by the existing 16 calculations

Appendix F

Workbook

Mathematics

Paper

Y6

Name:

Date:

School:

[redacted]

Calculate 456 + 372 =

Show your method

[redacted]

Calculate 20 x 46

Show your method

[redacted]

Calculate 35 x 99

Show your method

[redacted]

Calculate 500 - 76

Show your method

[redacted] **Calculate 401 ÷ 25**

Show your method

[redacted] **Calculate 5412 + 2584**

Show your method

[redacted] **Calculate 632 - 154**

Show your method

[redacted] **Calculate 27 x 63**

Show your method

[redacted]

Calculate 480 ÷ 20

Show your method

[redacted]

Calculate 245 + 256

Show your method

[redacted]

Calculate 568 x 34

Show your method

[redacted]

Calculate 693 ÷ 3

Show your method

[redacted]

Calculate 382 - 199

Show your method

[redacted]

Calculate 299 + 532

Show your method

[redacted]

Calculate 702 - 695

Show your method

[redacted]

Calculate 517 ÷ 19

Show your method

Appendix G

Confidence Questionnaire

Your maths questionnaire

Your name:………

Your school:……..

Question One: How good at maths are you?

Very good OK Not at all good

Question Two: If you were to list of all the pupils in your class from the worst to the best in maths, where would you put yourself? (Tick one box)

☐ *Very low* ☐ *Fairly low* ☐ *Near the middle* ☐ *Fairly high* ☐ *Very high*

Question Three: Some pupils are better at one subject than another. For example, you might be better at maths than reading. Compared to most of your other school subjects, how good are you at maths?

A lot better in maths than other subjects

About the same in maths than other subjects

A lot worse in maths than other subjects

Question Four: How well do you expect to do in maths this year?

Very well OK Not at all well

Question Five: How good would you be at learning something new in maths?

Very good OK Not at all good

Question Six: How hard is maths for you?

Not at all hard OK Very hard

Question Seven: Some things that you learn in school help you do things better out of the classroom. For example, learning about plants might help you to grow a garden. How useful is what you learn in maths?

Very useful OK Not at all useful

Question Eight: Being good at maths is

☐ *Not important* ☐ *Slightly important* ☐ *Important* ☐ *Quite important* ☐ *Very important*

Question Nine: In general, I find working on my maths lessons

Very interesting OK Very boring

Question Ten: How much do you like doing maths?

Very much It's OK Not at all

Appendix H

Script for class teacher

Workbooks and questionnaires not yet on desks

In today's lesson we are going to fill in the questionnaire and complete some calculations in a booklet that we talked about the other day. Your families were sent a letter about them. The questionnaire and workbook are not tests, but you must work by yourselves. So, first I am going to give each of you a questionnaire. When you get yours, please write your full name and our school at the top. Then, put your pencil down to show me that you are ready. Do not start to answers the questions.

Hand out questionniares

The questionnaire has ten questions. I will read aloud the questions one at a time. After I have read the question, please answer it and then put down your pencil. We will do the questions together, one at a time.

Read first question aloud, explaining that pupils should draw a circle around the face closest to their answer. When all pupils have finished, read the next question which requires pupils to tick a box. Then, work through the rest of the questions one at a time and retrieve the questionnaires before the next task. *The pilot questionnaires take around 10 minutes to complete, plus the time to hand them out and collect them back in again*

When you get your workbook, please write your full name and our school on the front cover. Then put your pencil down to show you are ready - please do not open it yet…wait until we are all ready.

Hand out workbooks

You will now have 16 questions to do by yourself. Try each question. If you cannot do a question, you can go on to the next one and then go back if you have time.

Has anyone got any questions they want to ask?

Respond to queries

You now have 40 minutes to do the questions, I will remind you when we are halfway through those 40 minutes, and when there are 5 minutes left.

Allow 40 minutes to complete.

Appendix I

Interview protocol

Date of interview:

Interviewee: Male/ Female   School Ref: Pupil Ref:

Start time:

End time:

Recorder ref:

	Interview protocol	Tick when completed			
1	Welcome pupil				
	Reinforce purpose of study				
	Confirm willingness to take part				
	Confirm pupil can stop at any point and return to class				
	Show pupil recorder and confirm audio only				
	Confirm pupil will receive a goodie bag at the end				
2	Show workbook and confirm it is their work				
	Begin with the photocopies of their four addition calculations Can you tell me why you used this way to work out your answer?	a	b	c	d
3	***Use follow up questions as required:*** Do you know another way/s? Why did you not use it/them? Did you check your answer? (If 'yes') How did you check it?	a	b	c	d
	Proceed to four subtraction calculations Can you tell me why you used this way to work out your answer?	a	b	c	d
	Use follow up questions as required: Do you know another ways? Why did you not use it/them? Did you check your answer? (If 'yes') How did you check it?	a	b	c	d
	Move on to four multiplication calculations Can you tell me why you used this way to work out your answer?	a	b	c	d
	Use follow up questions as required: Do you know another ways/? Why did you not use it/them? Did you check your answer? (If 'yes') How did you check it? a b c d	a	b	c	d
	Move on to four division calculations Can you tell me why you used this way to work out your answer?	a	b	c	d

	Use follow up questions as required: Do you know another way/s? Why did you not use it/them? Did you check your answer? (If 'yes') How did you check it? a b c d	a	b	c	d
	Ask pupil how important it is to get the right answer. Does speed matter?				
	Thank pupil for their help and give out goodie bag, return pupil to class				

Appendix J

Consent letter for parents and carers

Dear Parent or Carer

I am writing to ask your permission to allow your child to take part in a research study which will focus on the ways children complete their addition, subtraction, multiplication and division written calculations. Your child's school and class teacher have kindly agreed to co-operate with the project and now I am asking for your permission for your child to take part.

My name is Ems Lord and I am a second year PhD student with the University of Cambridge. I have designed a set of calculations which I would like to deliver to several classes of Year Six pupils. The findings from this research study will provide the data for my investigation into the written calculations of Year Six pupils.

I wish to deliver a short questionnaire about mathematics, followed by a booklet of 16 written calculations to your child's class during one of their daily mathematics lessons. The lesson would be led, as usual, by their teacher {insert teacher's name here}. Your child may also be invited to attend a short 20 minute interview, with myself, two weeks later to discuss the calculation choices that they made in their booklets. The interview would be audio recorded and conducted on school premises. Your child's participation would be voluntary and they would be reminded that they could return to their class at any time.

I would hope that your child's class would enjoy participating in a research project. Please be assured that I hold a full CRB check and I am also highly experienced classroom teacher. Your child's right to privacy will be rigorously upheld; the data would be made anonymous and your child's name would not appear on any computer record or memory stick, nor would it be revealed in any written report.

If you have any further questions about the project, please contact your child's head teacher. Should you wish to withdraw your child from the project, please complete and return the attached form to your child's class teacher by {state date}. Unless a signed form has been received by {state date}, it will be assumed with the agreement of the head teacher that your child has your permission to participate in this project.

Thank you for your help.

Ems Lord Signed:

…….

Your name:

I wish to withdraw my child …………………………………………… (name) from the research project.

Today's date:

Your signature:

Appendix K

Accuracy analysis for calculations (% in brackets)

456 + 372

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	519 (95.4)	276 (95.2)	243 (95.7)	25 (4.6)	14 (4.8)	11 (4.3)
Count- up	2 (66.7)	2 (66.7)	0 (0.0)	1 (33.3)	1 (33.3)	0 (0.0)
Expanded	6 (75.0)	4 (80.0)	2 (66.7)	2 (25.0)	1 (20.0)	1 (33.3)
Other	0 (0.0)	0 (0.0)	0 (0.0)	1 (100.0)	1 (100.0)	0 (0.0)
Partition	26 (76.5)	19 (70.4)	7 (77.8)	8 (23.5)	6 (29.6)	2 (22.2)
Total	553 (93.7)	301 (92.9)	252 (94.7)	37 (6.3)	23 (7.1)	14 (5.3)

5412 + 2584

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	521 (95.4)	279 (95.9)	242 (95.3)	25 (4.6)	12 (4.1)	13 (4.7)
Answer only	7 (87.5)	6 (85.7)	1 (100.0)	1 (12.5)	1 (14.3)	0 (0.0)
Count-up	1 (50.0)	1 (50.0)	0 (0.0)	1 (50.0)	1 (50.0)	0 (0.0)
Expanded	3 (100.0)	2 (100.0)	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	7 (100.0)	7 (100.0)	0 (100.0)
Partition	20 (83.3)	13 (86.7)	7 (77.8)	4 (16.7)	2 (13.3)	2 (22.2)
Total	552 (93.6)	301 (92.9)	251 (94.4)	38 (6.4)	23 (7.1)	15 (5.6)

245 + 256

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	485 (89.8)	270 (93.4)	215 (85.7)	55 (10.2)	19 (6.6)	36 (14.3)
Answer only	5 (83.3)	4 (80.0)	1 (100.0)	1 (16.7)	1 (20.0)	0 (0.0)
Count-up	1 (100.0)	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Doubling	1 (100.0)	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Expanded	7 (77.8)	4 (80.0)	3 (75.0)	2 (22.2)	1 (20.0)	1 (25.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	4 (100.0)	4 (100.0)	0 (100.0)
Partition	27 (93.1)	18 (94.7)	9 (90.0)	2 (6.9)	1 (5.3)	1 (10.0)
Total	526 (89.2)	298 (92.0)	228 (85.7)	64 (10.8)	26 (8.0)	38 (14.3)

299 + 532

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	464 (89.2)	250 (90.3)	214 (88.1)	56 (10.8)	27 (9.7)	29 (11.9)
Answer only	5 (83.3)	4 (80.0)	1 (100.0)	1 (16.7)	1 (20.0)	0 (0.0)
Count-up	1 (100.0)	0 (0.0)	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Expanded	4 (66.7)	3 (75.0)	1 (50.0)	2 (33.3)	1 (25.0)	1 (50.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	22 (100.0)	12 (100.0)	10 (100.0)
Partition	19 (86.4)	13 (81.3)	6 (100.0)	3 (13.6)	3 (18.7)	0 (0.0)
Round	10 (76.9)	8 (80.0)	2 (33.0)	3 (23.1)	2 (20.0)	1 (67.0)
Total	503 (85.3)	278 (85.8)	225 (84.6)	87 (14.7)	46 (14.2)	41 (15.4)

500 - 76

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	342 (79.5)	211 (87.9)	131 (68.9)	88 (20.5)	29 (12.1)	59 (31.1)
Answer only	8 (66.7)	2 (100.0)	6 (60.0)	4 (33.3)	0 (0.0)	4 (40.0)
Count back	18 (28.6)	0 (0.0)	18 (72.0)	45 (71.4)	38 (100.0)	7 (28.0)
Count-up	30 (56.6)	8 (50.0)	22 (88.0)	23 (43.4)	20 (50.0)	3 (12.0)
Expanded	2 (100.0)	1 (100.0)	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	6 (100.0)	4 (100.0)	2 (100.0)
Number bonds	13 (100.0)	8 (100.0)	5 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Other	0 (0.0)	0 (0.0)	0 (0.0)	1 (100.0)	0 (0.0)	1 (100.0)
Partition	3 (50.0)	2 (100.0)	1 (33.3)	3 (50.0)	0 (0.0)	3 (66.7)
Round	2 (66.7)	0 (0.0)	2 (66.7)	1 (33.3)	0 (100.0)	1 (33.3)
Same difference	1 (100.0)	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Total	419 (71.0)	233 (71.9)	186 (69.9)	171 (29.0)	91 (28.1)	80 (30.1)

632 -154

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	369 (73.5)	193 (71.0)	176 (76.5)	133 (26.5)	79 (29.0)	54 (23.5)
Answer only	5 (55.6)	4 (50.0)	1 (100.0)	4 (44.4)	4 (50.0)	0 (0.0)
Count back	14 (87.5)	9 (90.0)	5 (83.3)	2 (12.5)	1 (10.0)	1 (16.7)
Count-up	27 (77.1)	15 (83.3)	12 (70.6)	8 (22.9)	3 (16.7)	5 (29.4)
Expanded	1 (50.0)	0 (0.0)	1 (100.0)	1 (50.0)	1 (100.0)	0 (0.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	7 (100.0)	6 (100.0)	1 (100.0)
Other	1 (33.3)	1 (50.0)	0 (0.0)	2 (66.7)	1 (50.0)	1 (100.0)
Partition	2 (12.5)	1 (14.3)	1 (11.1)	14 (87.5)	6 (85.7)	8 (88.9)
Total	419 (71.0)	223 (68.8)	196 (73.7)	171 (29.0)	101 (31.2)	70 (26.3)

382 - 199

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	336 (71.0)	186 (73.2)	150 (68.5)	137 (29.0)	68 (26.8)	69 (31.5)
Answer only	2 (20.0)	2 (22.2)	0 (0.0)	8 (80.0)	7 (77.8)	1 (100.0)
Count back	6 (60.0)	4 (57.1)	2 (66.7)	4 (40.0)	3 (42.9)	1 (33.3)
Count-up	34 (89.5)	17 (100.0)	17 (81.0)	4 (10.5)	0 (0.0)	4 (19.0)
Expanded	0 (0.0)	0 (0.0)	0 (0.0)	1 (100.0)	1 (100.0)	0 (0.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	21 (100.0)	13 (100.0)	8 (100.0)
Other	1 (100.0)	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Partition	2 (8.3)	2 (25.0)	0 (0.0)	12 (91.7)	6 (75.0)	6 (100.0)
Round	20 (90.1)	12 (85.7)	8 (100.0)	2 (9.9)	2 (14.3)	0 (0.0)
Total	401 (68.0)	224 (69.1)	177 (66.5)	189 (32.0)	100 (30.9)	89 (33.5)

702 – 695

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	294 (66.2)	155 (66.5)	139 (65.9)	150 (33.8)	78 (33.5)	72 (34.1)
Answer only	14 (73.7)	12 (80.0)	2 (50.0)	5 (26.3)	3 (20.00	2 (50.0)
Count back	15 (75.0)	10 (71.4)	5 (83.3)	5 (25.0)	4 (28.6)	1 (75.0)
Count-up	57 (96.6)	30 (96.8)	27 (96.4)	2 (3.4)	1 (3.2)	1 (3.6)
Expanded	0 (0.0)	0 (0.0)	0 (0.0)	1 (100.0)	1 (100.0)	0 (0.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	31 (100.0)	2 (100.0)	11 (100.0)
Other	1 (33.3)	1 (100.0)	0 (0.0)	2 (66.7)	1 (50.0)	1 (50.0)
Partition	4 (30.8)	3 (37.5)	1 (20.0)	9 (69.2)	5 (62.5)	4 (80.0)
Total	385 (65.3)	211 (65.1)	174 (65.4)	205 (34.7)	113 (34.9)	92 (34.6)

20 x 46

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	144 (56.5)	76 (57.1)	68 (55.7)	111 (43.5)	57 (42.9)	54 (44.3)
Answer only	2 (33.3)	2 (50.0)	0 (0.0)	6 (66.7)	4 (50.0)	2 (100.0)
Count-up	3 (50.0)	3 (75.0)	0 (0.0)	3 (50.0)	1 (25.0)	2 (0.0)
Doubling	1 (50.0)	1 (100.0)	0 (0.0)	1 (50.0)	0 (0.0)	1 (100.0)
Expanded	27 (56.3)	21 (67.7)	6 (35.3)	21 (43.7)	10 (33.3)	11 (64.7)
Grid	121 (67.2)	69 (68.3)	52 (65.8)	59 (32.8)	32 (31.7)	27 (34.2)
Multiples of 10	24 (77.4)	12 (80.0)	12 (75.0)	7 (22.6)	3 (20.0)	4 (25.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	5 (100.0)	4 (100.0)	1 (100.0)
Other	2 (28.6)	2 (100.0)	0 (0.0)	5 (71.4)	1 (0.0)	4 (100.0)
Partition	24 (50.0)	15 (57.7)	9 (40.9)	24 (50.0)	11 (42.3)	13 (59.1)
Total	348 (59.0)	201 (62.0)	147 (55.3)	242 (41.0)	123 (38.0)	119 (44.7)

35 x 99

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	101 (40.7)	49 (37.7)	52 (44.1)	147 (59.3)	81 (62.3)	66 (55.9)
Answer only	1 (12.5)	1 (20.0)	0 (0.0)	7 (87.5)	4 (80.0)	3 (100.0)
Chunk up	0 (0.0)	0 (0.0)	0 (0.0)	2 (100.0)	2 (100.0)	0 (0.0)
Expanded	23 (50.0)	16 (59.3)	7 (36.8)	23 (50.0)	11 (41.7)	12 (63.2)
Grid	104 (52.3)	54 (48.6)	50 (56.8)	95 (47.7)	57 (51.4)	38 (43.2)
Multiples of 10	0 (0.0)	0 (0.0)	0 (0.0)	1 (100.0)	0 (0.0)	1 (100.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	7 (100.0)	6 (100.0)	1 (100.0)
Other	1 (11.1)	1 (50.0)	0 (0.0)	4 (89.9)	1 (50.0)	3 (100.0)
Partition	5 (10.9)	3 (12.0)	2 (9.5)	41 (89.1)	22 (88.0)	19 (90.5)
Round	18 (64.2)	12 (75.0)	6 (50.0)	10 (35.8)	4 (25.0)	6 (50.0)
Total	253 (42.9)	136 (42.0)	117 (44.0)	337 (57.1)	188 (58.0)	149 (56.0)

27 x 63

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	118 (45.7)	62 (45.3)	56 (46.3)	140 (54.3)	75 (54.7)	65 (53.7)
Grid	116 (62.7)	64 (62.7)	52 (61.9)	69 (37.3)	38 (37.3)	31 (38.1)
Expanded	26 (47.3)	13 (68.4)	13 (50.0)	29 (52.7)	16 (31.6)	13 (50.0)
Partition	5 (9.4)	5 (17.9)	0 (0.0)	48 (90.6)	23 (82.1)	25 (100.0)
Answer only	0 (0.0)	0 (0.0)	0 (0.0)	9 (100.0)	7 (100.0)	2 (100.0)
Other	3 (37.5)	1 (33.3)	2 (40.0)	5 (62.5)	2 (66.7)	3 (60.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	22 (100.0)	18 (100.0)	4 (100.0)
Total	268 (45.4)	145 (44.8)	123 (46.2)	322 (54.6)	179 (55.2)	143 (53.8)

568 x 34

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	91 (35.8)	52 (38.5)	39 (32.7)	163 (64.2)	83 (61.5)	80 (67.3)
Answer only	0 (0.0)	0 (0.0)	0 (0.0)	6 (100.0)	5 (100.0)	1 (100.0)
Count-up	3 (75.0)	0 (0.0)	3 (75.0)	1 (25.0)	0 (0.0)	1 (25.0)
Expanded	19 (42.2)	12 (46.2)	7 (36.8)	26 (57.8)	14 (53.8)	12 (63.2)
Grid	58 (32.2)	35 (35.7)	23 (28.0)	122 (67.8)	63 (64.3)	59 (72.0)
Multiples of 10	0 (0.0)	0 (0.0)	0 (0.0)	1 (100.0)	0 (0.0)	1 (100.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	52 (100.0)	32 (100.0)	20 (100.0)
Other	1 (14.3)	1 (25.0)	0 (0.0)	6 (85.7)	3 (75.0)	3 (100.0)
Partition	7 (17.1)	6 (25.0)	1 (6.3)	34 (82.9)	18 (75.0)	16 (93.7)
Total	179 (30.3)	106 (32.7)	73 (27.4)	411 (69.7)	218 (67.3)	193 (72.6)

401 ÷ 25

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	10 (43.7)	63 (47.4)	47 (39.5)	142 (56.3)	70 (52.6)	72 (60.5)
Answer only	22 (44.0)	16 (42.1)	6 (50.0)	28 (56.0)	22 (57.9)	6 (50.0)
Chunk down	50 (64.1)	28 (66.7)	22 (61.1)	28 (35.9)	14 (33.3)	14 (38.9)
Count-up	44 (65.6)	24 (64.9)	20 (66.7)	23 (34.4)	13 (35.1)	10 (33.3)
Grid	1 (14.3)	0 (0.0)	1 (33.3)	6 (85.7)	3 (100.0)	3 (66.7)
Multiples of 10	0 (0.0)	0 (0.0)	0 (0.0)	1 (100.0)	0 (0.0)	1 (100.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	94 (100.0)	57 (100.0)	37 (100.0)
Other	1 (3.1)	1 (9.1)	0 (0.0)	32 (96.9)	11 (90.9)	21 (100.0)
Partition	0 (0.0)	0 (0.0)	0 (0.0)	8 (100.0)	2 (100.0)	6 (100.0)
Total	28 (38.6)	132 (40.7)	96 (36.1)	362 (61.4)	192 (59.3)	170 (63.9)

480 ÷ 20

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	69 (66.3)	97 (71.9)	72 (60.0)	86 (33.7)	38 (28.1)	48 (40.0)
Answer only	30 (54.5)	25 (59.5)	5 (38.5)	25 (45.5)	17 (40.5)	8 (61.5)
Chunk down	46 (76.7)	22 (73.3)	24 (80.0)	14 (23.7)	8 (26.7)	6 (20.0)
Count-up	43 (66.2)	20 (62.5)	23 (69.7)	22 (33.8)	12 (37.5)	10 (30.3)
Grid	0 (0.0)	0 0.0)	0 (0.0)	3 (100.0)	2 (100.0)	1 (100.0)
Multiple of 10	4 (70.0)	8 (72.7)	6 (66.7)	6 (30.0)	3 (27.3)	3 (33.3)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	84 (100.0)	54 (100.0)	30 (100.0)
Other	1 (3.3)	0 (0.0)	1 (5.3)	29 (96.7)	11 (100.0)	18 (94.7)
Partition	6 (33.3)	4 (57.1)	2 (22.2)	12 (66.7)	3 (42.9)	9 (77.8)
Total	09 (52.4)	176 (54.3)	133 (50.0)	28 (47.6)	148 (45.7)	133 (50.0)

693 ÷ 3

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	287 (88.9)	156 (89.1)	131 (88.5)	36 (11.1)	19 (10.9)	17 (11.5)
Answer only	11 (30.6)	6 (31.6)	5 (29.4)	25 (69.4)	13 (68.4)	12 (70.4)
Chunk down	16 (55.2)	12 (60.0)	4 (36.4)	13 (44.8)	6 (40.0)	7 (63.6)
Count-up	8 (42.1)	3 (33.3)	5 (50.0)	11 (57.9)	6 (66.7)	5 (50.0)
Grid	1 (25.0)	1 (25.0)	0 (0.0)	3 (75.0)	3 (75.0)	0 (0.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	101 (100.0)	59 (100.0)	42 (100.0)
Other	3 (11.1)	1 (10.0)	2 (11.8)	24 (88.9)	9 (90.0)	15 (88.2)
Partition	25 (61.0)	15 (75.0)	10 (47.6)	16 (39.0)	5 (25.0)	11 (52.4)
Total	351 (59.5)	194 (59.9)	157 (59.0)	229 (40.5)	120 (40.1)	109 (41.0)

517 ÷ 19

	Number of correct answers			Number of incorrect answers		
Strategy	Overall	Boy	Girl	Overall	Boy	Girl
Formal algorithm	62 (26.6)	42 (33.1)	20 (18.9)	171 (73.4)	85 (66.9)	86 (81.1)
Answer only	0 (0.0)	0 (0.0)	0 (0.0)	33 (100.0)	22 (100.0)	11 (100.0)
Chunk down	48 (59.3)	26 (57.8)	22 (61.1)	33 (40.7)	19 (42.2)	14 (38.9)
Count-up	9 (28.1)	6 (31.6)	3 (23.1)	23 (71.9)	13 (68.4)	10 (76.9)
Expanded	2 (66.7)	2 (100.0)	0 (0.0)	1 (33.3)	0 (0.0)	1 (100.0)
Grid	1 (20.0)	0 (0.0)	1 (33.3)	4 (80.0)	2 (100.0)	2 (66.7)
Multiples of 10	0 (0.0)	0 (0.0)	0 (0.0)	1 (100.0)	0 (100.0)	1 (100.0)
No answer	0 (0.0)	0 (0.0)	0 (0.0)	146 (100.0)	84 (100.0)	62 (100.0)
Other	2 (5.0)	2 (13.3)	0 (0.0)	38 (95.0)	13 (86.7)	25 (100.0)
Partition	1 (5.3)	1 (12.5)	0 (0.0)	18 (94.7)	7 (87.5)	11 (100.0)
Total	123 (20.8)	77 (23.8)	46 (17.3)	467 (79.2)	247 (76.2)	220 (82.7)

Appendix L

Most Accurate strategies: One-way ANOVA tables and post-hoc Tukey tests

456+372

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	2.210911	3	0.73697	13.30093	2.07E-08	3.002737
Within Groups	32.46875	586	0.055407			
Total	34.67966	589				

Note. F-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Expanded vs other	2.4528	0.3069402
Expanded vs partition	0.2248	0.8999947
Expanded vs formal algorithm	3.4422	0.0720480
Other vs partition	3.0087	0.1455799
Other vs formal algorithm	5..4359	0.0010053**
Partition vs formal algorithm	6.4349	0.0010053**

Note. ** p<0.01, * p<0.05.

5412+2584

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	7.158389	4	1.789597	35.77526	9.31E-27	2.698
Within Groups	29.26364	585	0.050023			
Total	36.42203	589				

Note. F-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs other	3.0501	0.1977832
Answer only vs no answer	10.6902	0.0010053**
Answer only vs partition	0.6453	0.8999947
Answer only vs formal algorithm	1.4064	0.8408407
Other vs no answer	6.4792	0.0010053**
Other vs partition	3.0012	0.2120004
Other vs formal algorithm	4.9854	0.0041555**
No answer vs partition	12.2665	0.0010053**
No answer vs formal algorithm	15.8620	0.0010053**
Partition vs formal algorithm	3.6648	0.0734161

Note. ** $p<0.01$, * $p<0.05$.

245 + 256

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	1.337664	4	0.334416	3.561165	0.006989	2.698
Within Groups	54.93522	585	0.093906			
Total	56.27288	589				

Note. F-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	Tukey *p*-value
Answer only vs other	2.4286	0.4257715
Answer only vs expanded	0.6071	0.8999947
Answer only vs partition	1.2488	0.8999947
Answer only vs formal algorithm	1.0082	0.8999947
Other vs expanded	3.4612	0.1044211
Other vs partition	4.9811	0.0041998**
Other vs formal algorithm	5.1590	0.0026584**
Expanded vs partition	0.6475	0.8999947
Expanded vs formal algorithm	0.2999	0.8999947
Partition vs formal algorithm	0.7962	0.8999947

Note. ** $p<0.01$, * $p<0.05$.

299 + 532

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	17.13669	6	2.856115	29.19487	1.29E-30	2.347921
Within Groups	57.0345	583	0.097829			
Total	74.17119	589				

Note. F-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs expanded	1.3052	0.8999947
Answer only vs no answer	8.1810	0.0010053**
Answer only vs partition	0.2975	0.8999947
Answer only vs round	0.5873	0.8999947
Answer only vs formal algorithm	0.6494	0.8999947
Expanded vs no answer	6.5448	0.0010053**
Expanded vs partition	1.9337	0.7192459
Expanded vs round	0.9396	0.8999947
Expanded vs formal algorithm	2.4848	0.4944882
No answer vs partition	12.9511	0.0010053**
No answer vs round	9.9423	0.0010053**
No answer vs formal algorithm	18.5357	0.0010053**
Partition vs round	1.2202	0.8999947
Partition vs formal algorithm	0.5956	0.8999947
Round vs formal algorithm	1.9818	0.6996411

Note. ** $p<0.01$, * $p<0.05$. ‘Other’ category excluded from Tukey test due to insufficient data.

500 – 76

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	7.28441	7	1.04063	5.386106	5.37E-06	2.235671
Within Groups	112.4461	582	0.193206			
Total	119.7305	589				

Note. F-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs count back	1.9119	0.8703824
Answer only vs count-up	4.4066	0.0402833*
Answer only vs other	0.4386	0.8999947
Answer only vs no answer	4.8473	0.0149669*
Answer only vs number bonds	2.6740	0.5495264
Answer only vs partition	1.1670	0.8999947
Answer only vs formal algorithm	0.6326	0.8999947
Count back vs count-up	2.8589	0.4699527
Count back vs other	0.5127	0.8999947
Count back vs no answer	5.8571	0.0010275**
Count back vs number bonds	1.5346	0.8999947
Count back vs partition	2.0918	0.7946495
Count back vs formal algorithm	2.0655	0.8057187
Count-up vs other	1.8330	0.8999947
Count-up vs no answer	7.0467	0.0010053**
Count-up vs number bonds	0.2112	0.8999947
Count-up vs partition	3.3120	0.2727789
Count-up vs formal algorithm	5.5611	0.0023845**
Other vs no answer	4.1308	0.0705719
Other vs number bonds	1.4329	0.8999947
Other vs partition	1.2392	0.8999947
Other vs formal algorithm	0.1984	0.8999947
No answer vs number bonds	6.0175	0.0010053**
No answer vs partition	2.7863	0.5022293
No answer vs formal algorithm	5.4037	0.0036601**
Number bonds vs partition	2.7580	0.5141624
Number bonds vs formal algorithm	2.6522	0.5587134
Partition vs formal algorithm	1.4924	0.8999947

Note. ** $p<0.01$, * $p<0.05$.

632 – 154

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	10.58238	6	1.763731	9.275541	1E-09	2.347921
Within Groups	110.8566	583	0.190149			
Total	121.439	589				

Note. *F*-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs count back	2.4864	0.5661034
Answer only vs count-up	1.8733	0.8201825
Answer only vs other	0.9045	0.8999947
Answer only vs no answer	3.5752	0.1511127
Answer only vs partition	3.3513	0.2134069
Answer only vs formal algorithm	1.7310	0.8791131
Count back vs count-up	1.1131	0.8999947
Count back vs other	3.0068	0.3387403
Count back vs no answer	6.2621	0.0010053**
Count back vs partition	6.8798	0.0010053**
Count back vs formal algorithm	1.7871	0.8558648
Count-up vs other	2.5196	0.5523519
Count-up vs no answer	6.0426	0.0010053**
Count-up vs partition	6.9470	0.0010053**
Count-up vs formal algorithm	0.6747	0.8999947
Other vs no answer	2.2155	0.6783637
Other vs partition	1.7408	0.8750892
Other vs formal algorithm	2.4178	0.5945315
No answer vs partition	0.8946	0.8999947
No answer vs formal algorithm	6.2637	0.0010053**
Partition vs formal algorithm	7.7909	0.0010053**

Note. ** $p<0.01$, * $p<0.05$.

382 – 199

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	19.52528	7	2.789325	14.90294	6.07E-18	2.235671
Within Groups	108.9307	582	0.187166			
Total	128.4559	589				

Note. *F*-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs count back	2.9238	0.4400920
Answer only vs count-up	6.3899	0.0010053**
Answer only vs other	1.2660	0.8999947
Answer only vs no answer	1.7016	0.8999947
Answer only vs partition	0.4512	0.8999947
Answer only vs round	6.0777	0.0010053**
Answer only vs formal algorithm	5.2208	0.0059236**
Count back vs count-up	2.7109	0.5340085
Count back vs other	0.4220	0.8999947
Count back vs no answer	5.1048	0.0079650**
Count back vs partition	3.6092	0.1757926
Count back vs round	2.6493	0.5599468
Count back vs formal algorithm	1.1289	0.8999947
Count-up vs other	1.7786	0.8999947
Count-up vs no answer	10.7565	0.0010053**
Count-up vs partition	7.8615	0.0010053**
Count-up vs round	0.1751	0.8999947
Count-up vs formal algorithm	3.5745	0.1856632
Other vs no answer	2.2087	0.7454479
Other vs partition	1.5444	0.8999947
Other vs round	1.8107	0.8999947
Other vs formal algorithm	0.9704	0.8999947
No answer vs partition	1.3535	0.8999947
No answer vs round	9.7408	0.0010053**
No answer vs formal algorithm	10.4125	0.0010053**
Partition vs round	7.3263	0.0010053**
Partition vs formal algorithm	6.8407	0.0010053**
Round vs formal algorithm	2.9786	0.4146190

Note. ** $p<0.01$, * $p<0.05$.

702-695

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	21.56122	6	3.593536	18.67064	6.49E-20	2.347921
Within Groups	112.21	583	0.19247			
Total	133.7712	589				

Note. F-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs count back	0.1324	0.8999947
Answer only vs count-up	2.8017	0.4301796
Answer only vs other	2.8528	0.4068853
Answer only vs no answer	8.1523	0.0010053**
Answer only vs partition	3.8434	0.0955135
Answer only vs formal algorithm	1.0276	0.8999947
Count back vs count-up	2.6923	0.4796306
Count back vs other	2.9427	0.3664884
Count back vs no answer	8.4296	0.0010053**
Count back vs partition	4.0021	0.0716043
Count back vs formal algorithm	1.2387	0.8999947
Count-up vs other	4.4678	0.0275856*
Count-up vs no answer	14.0392	0.0010053**
Count-up vs partition	6.9273	0.0010053**
Count-up vs formal algorithm	7.0706	0.0010053**
Other vs no answer	1.5169	0.8999947
Other vs partition	0.3253	0.8999947
Other vs formal algorithm	2.6454	0.5002460
No answer vs partition	3.0018	0.3408573
No answer vs formal algorithm	11.4901	0.0010053**
Partition vs formal algorithm	4.0609	0.0639262

Note. ** $p<0.01$, * $p<0.05$.

20 x 46

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	6.163664	8	0.770458	3.277577	0.001146	2.146541
Within Groups	136.5753	581	0.235069			
Total	142.739	589				

Note. *F*-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs count-up	1.3502	0.8999947
Answer only vs other	0.5002	0.8999947
Answer only vs expanded	2.3869	0.7262173
Answer only vs grid	3.4085	0.2804286
Answer only vs multiples of 10	3.8557	0.1405443
Answer only vs no answer	1.2791	0.8999947
Answer only vs partition	1.9095	0.8999947
Answer only vs formal algorithm	2.5566	0.6537587
Count-up vs other	0.9224	0.8999947
Count-up vs expanded	0.4210	0.8999947
Count-up vs grid	1.2105	0.8999947
Count-up vs multiples of 10	1.7932	0.8999947
Count-up vs no answer	2.4085	0.7169941
Count-up vs partition	0.0000	0.8999947
Count-up vs formal algorithm	0.4570	0.8999947
Other vs expanded	1.8402	0.8999947
Other vs grid	2.8940	0.5096385
Other vs multiples of 10	3.3962	0.2852053
Other vs no answer	1.7432	0.8999947
Other vs partition	1.3384	0.8999947
Other vs formal algorithm	1.9898	0.8958126
Expanded vs grid	1.9702	0.8999947
Expanded vs multiples of 10	2.6799	0.6011052
Expanded vs no answer	3.4915	0.2493523
Expanded vs partition	0.8931	0.8999947
Expanded vs formal algoruthm	0.0409	0.8999947
Grid vs multiples of 10	1.5296	0.8999947
Grid vs no answer	4.3248	0.0588358
Grid vs partition	3.0924	0.4186880
Grid vs formal algorithm	3.2215	0.3583119
Multiples of 10 vs no answer	4.6858	0.0271275*
Multiples of 10 vs partition	3.4711	0.2567926
Multiples of 10 vs formal algorithm	3.2125	0.3624061
No answer vs partition	3.1035	0.4134193
No answer vs formal algorithm	3.6476	0.1974911
Partition vs formal algorithm	1.1996	0.8999947

Note. ** $p<0.01$, * $p<0.05$.

35 x 99

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	10.8599	7	1.551414	6.755864	1.02E-07	2.235671
Within Groups	133.6503	582	0.22964			
Total	144.5102	589				

Note. *F*-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs other	2.8890	0.4561570
Answer only vs expanded	0.0000	0.8999947
Answer only vs grid	3.2542	0.2948599
Answer only vs no answer	0.7128	0.8999947
Answer only vs partition	0.1256	0.8999947
Answer only vs round	3.8122	0.1254461
Answer only vs formal algorithm	2.3189	0.6990241
Other vs expanded	2.8890	0.4561570
Other vs grid	0.4079	0.8999947
Other vs no answer	3.6371	0.1686058
Other vs partition	5.5382	0.0025387**
Other vs round	1.7589	0.8999947
Other vs formal algorithm	1.7049	0.8999947
Expanded vs grid	3.2542	0.2948599
Expanded vs no answer	0.7128	0.8999947
Expanded vs partition	0.1256	0.8999947
Expanded vs round	3.8122	0.1254461
Expanded vs formal algorithm	2.3189	0.6990241
Grid vs no answer	4.0106	0.0879973
Grid vs partition	7.4667	0.0010053**
Grid vs round	1.7581	0.8999947
Grid vs formal algorithm	3.5771	0.1849343
No answer vs partition	0.7907	0.8999947
No answer vs round	4.4895	0.0337314*
No answer vs formal algorithm	3.1359	0.3430475
Partition vs round	6.5767	0.0010053**
Partition vs formal algorithm	5.4886	0.0029068**
Round vs formal algorithm	3.4875	0.2122347

Note. ** $p<0.01$, * $p<0.05$.

27 x 63

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	18.86458	6	3.144096	14.3777	2.53E-15	2.347921
Within Groups	127.4897	583	0.218679			
Total	146.3542	589				

Note. *F*-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs expanded	3.9759	0.0752514
Answer only vs other	2.3339	0.6292974
Answer only vs grid	5.5553	0.0018446**
Answer only vs no answer	0.0000	0.8999947
Answer only vs partition	0.7914	0.8999947
Answer only vs formal algorithm	4.1135	0.0576253
Expanded vs other	0.7811	0.8999947
Expanded vs grid	3.0384	0.3255008
Expanded vs no answer	5.6672	0.0013459**
Expanded vs partition	5.9451	0.0010053**
Expanded vs formal algorithm	0.2339	0.8999947
Other vs grid	2.1106	0.7218201
Other vs no answer	2.7469	0.4551123
Other vs partition	2.2377	0.6691480
Other vs formal algorithm	0.7265	0.8999947
Grid vs no answer	8.4083	0.0010053**
Grid vs partition	10.3400	0.0010053**
Grid vs formal algorithm	5.2042	0.0047379**
No answer vs partition	1.1249	0.8999947
No answer vs formal algorithm	6.2803	0.0010053**
Partition vs formal algorithm	7.3575	0.0010053**

Note. ** $p<0.01$, * $p<0.05$.

568 x 34

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	7.535149	6	1.255858	6.24938	2.24E-06	2.347921
Within Groups	117.1581	583	0.200957			
Total	124.6932	589				

Note. F-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs expanded	3.0648	0.3150953
Answer only vs grid	2.4495	0.5814092
Answer only vs other	2.1032	0.7249175
Answer only vs no answer	0.0000	0.8999947
Answer only vs partition	1.2322	0.8999947
Answer only vs formal algorithm	2.7364	0.4598570
Expanded vs grid	1.8928	0.8120641
Expanded vs other	0.8631	0.8999947
Expanded vs no answer	6.5422	0.0010053**
Expanded vs partition	3.6748	0.1278419
Expanded vs formal algorithm	1.2474	0.8999947
Grid vs other	0.1176	0.8999947
Grid vs no answer	6.4567	0.0010053**
Grid vs partition	2.7617	0.4483771
Grid vs formal algorithm	1.1671	0.8999947
Other vs no answer	3.2836	0.2352609
Other vs partition	1.5629	0.8999947
Other vs formal algorithm	0.2663	0.8999947
No answer vs partition	2.5789	0.5277970
No answer vs formal algorithm	7.4256	0.0010053**
Partition vs formal algorithm	3.5152	0.1665154

Note. ** $p<0.01$, * $p<0.05$.

401 ÷ 25

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	28.86964	7	4.124235	21.5763	6.1E-26	2.235671
Within Groups	111.2473	582	0.191147			
Total	140.1169	589				

Note. *F*-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs chunk down	4.1755	0.0646251
Answer only vs count-up	1.3449	0.8999947
Answer only vs grid	2.3817	0.6725847
Answer only vs other	5.9748	0.0010053**
Answer only vs no answer	7.7380	0.0010053**
Answer only vs partition	3.7377	0.1426375
Answer only vs formal algorithm	0.0730	0.8999947
Chunk down vs count-up	1.4237	0.8999947
Chunk down vs grid	4.2735	0.0530172
Chunk down vs other	10.4788	0.0010053**
Chunk down vs no answer	15.0388	0.0010053**
Chunk down vs partition	5.8185	0.0011499**
Chunk down vs formal algorithm	6.4899	0.0010053**
Count-up vs grid	2.9989	0.4051753
Count-up vs other	5.9756	0.0010053**
Count-up vs no answer	6.9452	0.0010053**
Count-up vs partition	4.2528	0.0553101
Count-up vs formal algorithm	1.5803	0.8999947
Grid vs other	0.8841	0.8999947
Grid vs no answer	1.0038	0.8999947
Grid vs partition	0.8929	0.8999947
Grid vs formal algorithm	2.4789	0.6316674
Other vs no answer	0.1315	0.8999947
Other vs partition	0.2421	0.8999947
Other vs formal algorithm	7.2075	0.0010053**
No answer vs partition	0.1869	0.8999947
No answer vs formal algorithm	11.1134	0.0010053**
Partition vs formal algorithm	3.9317	0.1019777

Note. ** $p<0.01$, * $p<0.05$.

480 ÷ 20

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	41.26056	7	5.894366	32.39175	4.94E-38	2.235671
Within Groups	105.9072	582	0.181971			
Total	147.1678	589				

Note. F-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs chunk down	3.9285	0.1025802
Answer only vs count-up	2.1006	0.7909727
Answer only vs multiples of 10	1.7287	0.8999947
Answer only vs no answer	10.4252	0.0010053**
Answer only vs other	7.3950	0.0010053**
Answer only vs partition	2.5897	0.5850127
Answer only vs formal algorithm	2.6155	0.5741787
Chunk down vs count-up	1.9468	0.8557292
Chunk down vs multiples of 10	1.0384	0.8999947
Chunk down vs no answer	15.0368	0.0010053**
Chunk down vs other	10.9321	0.0010053**
Chunk down vs partition	5.3456	0.0042730**
Chunk down vs formal algorithm	2.4011	0.6644312
Count-up vs multiples of 10	0.2882	0.8999947
Count-up vs no answer	13.2761	0.0010053**
Count-up vs other	9.4407	0.0010053**
Count-up vs partition	4.0852	0.0770536
Count-up vs formal algorithm	0.0288	0.8999947
Multiples of 10 vs no answer	8.9290	0.0010053**
Multiples of 10 vs other	7.2384	0.0010053**
Multiples of 10 vs partition	3.5366	0.1969587
Multiples of 10 vs formal algorithm	0.2992	0.8999947
No answer vs other	0.9594	0.8999947
No answer vs partition	4.2547	0.0550979
No answer vs formal algorithm	17.4651	0.0010053**
Other vs partition	3.1221	0.3491185
Other vs formal algorithm	10.9662	0.0010053**
Partition vs formal algorithm	4.4779	0.0345886*

Note. ** $p<0.01$, * $p<0.05$.

693 ÷ 3

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	72.44382	6	12.07397	104.0165	8.45E-89	2.347921
Within Groups	67.67313	583	0.116077			
Total	140.1169	589				

Note. F-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs chunk down	1.6666	0.8999947
Answer only vs count-up	0.5399	0.8999947
Answer only vs other	6.3061	0.0010053**
Answer only vs no answer	10.6533	0.0010053**
Answer only vs partition	2.9615	0.3582362
Answer only vs formal algorithm	11.3905	0.0010053**
Chunk down vs count-up	1.8377	0.8349086
Chunk down vs other	7.2003	0.0010053**
Chunk down vs no answer	10.8705	0.001005**3
Chunk down vs partition	0.9928	0.8999947
Chunk down vs formal algorithm	7.2205	0.0010053**
Count-up vs other	4.5452	0.0232505*
Count-up vs no answer	6.9892	0.0010053**
Count-up vs partition	2.8224	0.4207337
Count-up vs formal algorithm	8.2269	0.0010053**
Other vs no answer	1.9963	0.7691944
Other vs partition	8.8070	0.0010053**
Other vs formal algorithm	17.1589	0.0010053**
No answer vs partition	13.6680	0.0010053**
No answer vs formal algorithm	32.3763	0.0010053**
Partition vs formal algorithm	6.9899	0.0010053**

Note. ** $p<0.01$, * $p<0.05$.

517 ÷ 19

ANOVA

Source	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	21.63165	7	3.090236	23.56939	3E-28	2.235671
Within Groups	76.30733	582	0.131112			
Total	97.93898	589				

Note. *F*-value exceeded critical value hence post-hoc Tukey tests conducted on each pair.

Post-hoc Tukey

Treatment pair	Tukey Q statistic	*p*-value
Answer only vs chunk down	11.2072	0.0010053**
Answer only vs count-up	4.4275	0.0385351*
Answer only vs grid	1.6277	0.8999947
Answer only vs no answer	0.1397	0.8999947
Answer only vs other	0.8304	0.8999947
Answer only vs partition	0.7138	0.8999947
Answer only vs formal algorithm	5.5429	0.0025069**
Chunk down vs count-up	5.8238	0.0011330**
Chunk down vs grid	3.3275	0.2669990
Chunk down vs no answer	16.4907	0.0010053**
Chunk down vs other	10.9660	0.0010053**
Chunk down vs partition	8.2732	0.0010053**
Chunk down vs formal algorithm	9.9657	0.0010053**
Count-up vs grid	0.6599	0.8999947
Count-up vs no answer	5.4863	0.0029250**
Count-up vs other	3.8081	0.1263241
Count-up vs partition	3.0830	0.3665774
Count-up vs formal algorithm	0.3611	0.8999947
Grid vs no answer	1.6581	0.8999947
Grid vs other	1.2351	0.8999947
Grid vs partition	1.1451	0.8999947
Grid vs formal algorithm	0.5516	0.8999947
No answer vs other	0.9426	0.8999947
No answer vs partition	0.7321	0.8999947
No answer vs formal algorithm	9.5025	0.0010053**
Other vs partition	0.0369	0.8999947
Other vs formal algorithm	4.8827	0.0137560**
Partition vs formal algorithm	3.4584	0.2216349

Note. ** p<0.01, * p<0.05.

Appendix M

Levels of Correlation Between the Predictor Variables

	KS1 Level	Confidence	Gender
KS1 Level	1.00	0.31	-0.01
Confidence	0.31	1.00	-0.13
Gender	-0.01	-0.13	1.00