Explaining Individual Differences in Strategy Variability Amongst Secondary School Mathematics Students

By Tim M. H. Jay, BSc.

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy, May 2007

Abstract

This thesis reports an investigation of individual differences in children's learning of a concept in mathematics, involving rates of change on linear graphs. Evidence in the literature suggests both that high levels of strategy variability are associated with conceptual change in mathematics and that there are individual differences in strategy variability. Therefore it is argued that differences in strategy variability can offer useful insight into children's learning of mathematics.

A series of experiments are reported that each aimed to explore individual and group differences in strategy variability amongst secondary school mathematics students. Methods used for data collection progressed from whole-class testing of students, to individual testing, to individual interviews employing think-aloud protocols, as the need grew for increasingly detailed data on children's strategies for solving problems. Early studies showed a gender difference in strategy variability, so later studies were designed to elaborate on and clarify this relationship. In combination, the results of the studies reported here suggest that there are robust differences in strategy variability between boys and girls and that this effect interacts with the context in which the problems are solved. The use of think-aloud protocols produced a complete reversal of the gender effect on strategy variability.

The implications of these findings are discussed, both in terms of learning theory and in terms of their potential impact on the mathematics classroom. The main contribution of this investigation to the literature is in helping to establish strategy variability as a key to understanding cognitive development and as an indicator of children's specific needs for intervention and support in the classroom.

Acknowledgements

There are a number of people without whose support this thesis would not have been written.

My supervisors, Professor David Wood, Dr. Peter Gates and Dr. Shaaron Ainsworth have provided high levels of enthusiasm, support and guidance between them. I feel privileged to have worked with each of them.

I would like to thank the schools, teachers and students that have participated in the research presented here. I have been constantly surprised by the levels of enthusiasm and support for my research shown to me by participating schools. Without them, there would be no research to report.

The examiners for this thesis provided some very thoughtful comments that led to the development of the final version.

Also, obviously, thanks to Claire and Eddy – I don't think I need to say why.

List of tables

List of figures

Contents

1 Introduction

1.1 Summary of the thesis

A key aspect of children's learning is the way in which the application of one strategy from a range of strategies to a given problem changes over time. It is becoming clear that children's strategy choices in response to a particular problem set do not involve a simple succession of stable states, but feature periods of variability in which strategy choices vary from trial to trial. This thesis describes an investigation of individual differences in strategy variability in response to mathematical problems. The main focus of the investigation involves an attempt to account for individual differences in strategy variability by identifying one or more group factors associated with variability.

The literature review in section 2 describes the state of research at present in relation to strategy development and explains the need for further development in research pertaining to variability. The evidence for individual differences in strategy variability is examined, as are candidate group factors for an association with variability. Also in section 2 is some discussion of the domain under investigation in this thesis – rate of change problems – and the reasons for the use of this domain in this investigation. There will also be consideration of the theoretical framework in which the research to be reported here has been conducted, and a description of the methodological standpoint.

In section 3, there is a report of the first study conducted as part of this investigation. In this study, gender is identified as a group factor that can potentially explain a considerable portion of the variance in strategy variability. The remainder of the thesis following section 3 has the aim of clarifying and elaborating on the relationship between gender and strategy variability.

Section 4 contains a review of the research literature on gender differences in mathematics, including research on achievement and participation as well as research on aspects of gender more pertinent to the investigation presented here involving gender differences in affective and cognitive factors involved in mathematical reasoning and gender differences in strategy use in response to mathematical problems.

Section 5 describes the second study in this investigation, in which confidence is gained in the association between gender and strategy variability. Also in this section, there is some discussion regarding the need for a more fine-grained analysis of the gender effect. To this end, section 6 contains some discussion of possible ways to obtain the data required for this deeper analysis.

In section 7, a third study is described, in which think-aloud protocols and retrospective reports are used in order to analyse the gender effect on variability in greater detail. There is some discussion in this section concerning the effect of the use of think-alouds on children's strategy choices across trials. Section 8 describes a follow-up, fourth study, in which the sample of participants from study three is revisited in order to test the status of strategy variability as a stable characteristic of children's mathematical thinking. The final study reported in this thesis is found in section 9, and aims to show the range of the gender effect in terms of children's level of achievement in relation to the problem space.

Finally, in section 10, there is a discussion of various aspects of the investigation , including an assessment of the gender effect of strategy variability, children's reactivity to think-aloud protocols and potential implications of these findings for both future research in this area and for the classroom.

1.2 Background

The purpose of this section is to describe the problems that I was considering before the beginning of my PhD studies and why I think they are important. The aim of this section is to provide the context for the rest of this thesis.

Before embarking on my research I had spent over two years teaching mathematics in a secondary school in Devon, England. Before this, I had completed a PGCE in Secondary Mathematics at the University of Nottingham. My decision to spend some time teaching was due to my interests in the psychology of learning generated during undergraduate studies in Psychology, also at the University of Nottingham. My reasoning was that teaching was an ideal way to get practical insight into the ways in which children learn.

The school that I worked at, South Dartmoor Community College, was a true comprehensive school. Being located on the edge of Dartmoor in Devon, the catchment area spanned over 100 square miles, with children being bussed in from various villages and small towns. There was a wide distribution of mathematical ability amongst students. The mathematics department set children according to ability, with set 1 the most able and set 5 the least – I taught a number of groups at either end of that scale. Some children that I worked with would go on to achieve the highest grade at GCSE (school-leaving exam in England and Wales, taken at age 16) with little apparent effort, whilst some would struggle to achieve even a pass.

I understood at that time that there were many factors involved in students' level of performance in school mathematics. I knew that children came into my classes from a wide variety of backgrounds and that this had some association with their ability in mathematics. Some of my students had parents that could provide a stable environment and offer help with homework and coursework, while some had a very difficult home life and various issues to overcome at home before even thinking about schoolwork. Many of these factors seemed to be difficult to integrate into my classroom practice. It was my belief that when children came into my class, there were so many such factors (socio-economic class, parental support, amount of sleep, primary school background, number of siblings, amount of private time and space – there is probably not sufficient space for an exhaustive list even in a PhD thesis) that I decided to simply accept the fact that I was not in a position to deal with them all. It seems to me that it is not the classroom teacher's responsibility to deal with this wide range of social factors. Even if I understood their effects, I was not in a position to remedy any of them – all I could do was accept their existence and their detrimental effects on my students. This is not to say that I do not value research in the sociology of mathematics education. I believe that such research is unquestionably important, just on a much wider scale than the individual classroom teacher can possibly be concerned with.

Within the classroom, I decided to limit myself to the task of trying to understand and support the cognitive processes by which children learned new mathematical concepts. With time, of course, I realized that this was no smaller a task than the one that I had rejected.

Over two years, I was at least able to focus a little more closely on the kinds of questions I was interested in. I was particularly interested in difference. The main thing that I learnt during my short teaching career was not to think of a class of children as a homogeneous group. There seemed to be huge differences within each group of children in the ways that children constructed concepts about mathematical objects. This is where I decided to focus my attention. It seemed to me that an interesting thing to look at was the way in which children changed from one way of solving a mathematical problem to another. I enjoyed following children's progress from the use of inappropriate and/or inefficient methods for solving problems to the use of appropriate and efficient methods for those same problems. Any differences in the ways that children made that transition between such methods seemed to have the potential to explain performance in mathematics more generally. Understanding of these cognitive differences seemed also to offer the potential to be useful for the classroom teacher in a way that understanding of social factors could not. I felt that if I could understand the differences in the ways that children constructed their understanding of mathematics, then I would be in a better position to support the accumulation of those understandings through appropriate support and intervention. There is nothing wrong with a little idealism!

That is the position towards which I have, during the last three years, been helping clear a path. I have aimed to investigate the ways in which children differ in terms of their development of mathematical understanding, intending to always keep in mind the ways in which my research will be useful for teachers and learners of mathematics.

2 Literature Review

2.1 Variation in Achievement in School Mathematics

There is a high degree of variation amongst children in terms of achievement in mathematics. This section will describe the extent of that variation and help explain the need for further investigation of its sources.

Figure 1 shows the distribution of children across National Curriculum levels shown by children in the 2005 Key Stage 3 Mathematics SAT (DFES, 2005b). The Key Stage 3 test is taken by children at age 14 and is intended to be a general measure of children's mathematics achievement used for various purposes including the determination of children's point of entry to the GCSE curriculum and the assessment of school performance for the publication of league tables. To help illustrate the range of children's levels of achievement at 14, Table 1 gives some descriptors of each level. It is clear from the graph that within a single year group there is substantial variation in children's achievement. A year 9 group in a typical school may well contain children with achievements ranging from that of the average level of a 7-year-old right up to the average level of the top 5% of 16-year-olds.

Similar data can be found in Shayer, Kuchemann and Wylam (1976) and Shayer and Wylam (1978), where children's achievement on Piagetian tests again shows wide variation in the mental development shown by any one year group.

Figure 1: Distribution of children across National Curriculum Levels in UK Key Stage 3 tests 2004/5

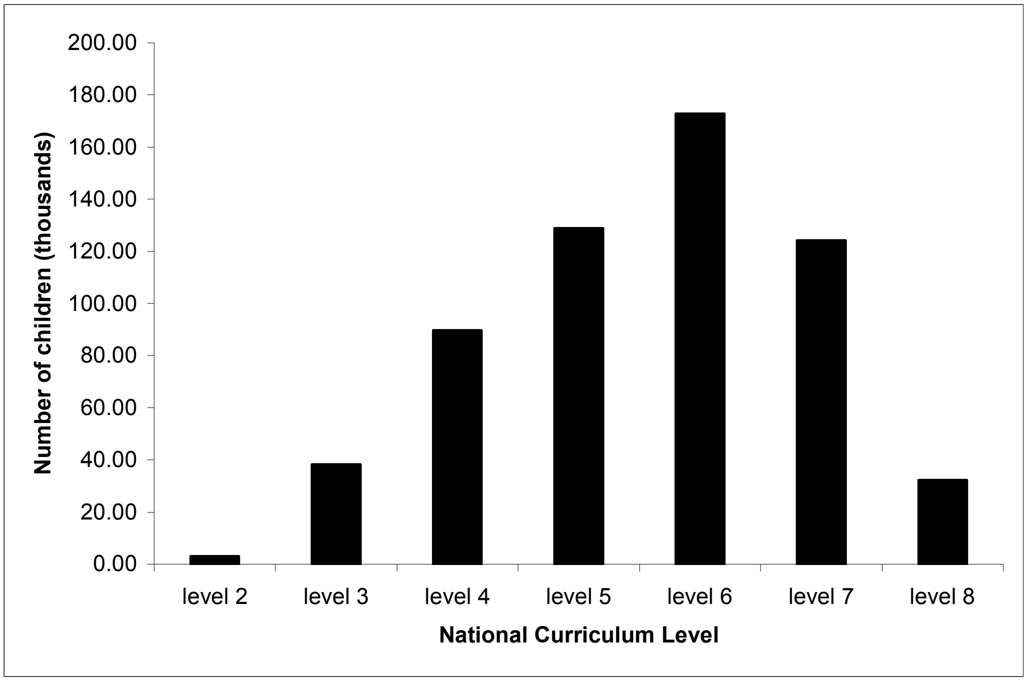


Table 1: Descriptors of National Curriculum levels

National Curriculum Level	Descriptor of level
2	Level of average 7-year-old
3	Level of average 8-year old
4	Level of average 11-year-old
5	Level of average 14-year-old
6	Level of average 16-year-old
7	Level of top 15% of 16-year-olds
8	Level of top 5% of 16-year-olds

These analyses were based on the CSMS survey (Hart, 1981) and show that children's level of mathematical development at 14 years-old, for example, assessed using Piagetian levels, ranges from the average level of a 6-year old up to the average level of the top 15% of 16 year-olds. This means that approximately half of this age group have reached Piaget's formal operational stage, while half remain in the concrete operational stage (see Piaget & Inhelder, 1969).

The data presented above are not intended to imply that the Mathematics classroom holds a special position in terms of the variability in achievement that can be observed. Similar distributions of achievement can be observed in Key Stage 3 results for English and Science (although the distribution is broader and flatter for Maths than English and Science , see DFES (2005b) – plus level 8 is accessible only in Mathematics – it is not possible to separate ability from the design of the test). Neither does this distribution reflect the range of ability levels that might be observed within a single classroom, or even necessarily within a school, due to setting or streaming according to ability. The intent of the discussion presented here is, rather, to argue that variance in achievement in Key Stage 3 mathematics is so large as to demand explanation of its origin and further, that this origin may in part lie in fundamental differences in the ways that individuals develop new concepts in Mathematics (or English or Science… However, we will only be concerned with Mathematics here!).

The range of achievement levels in Mathematics classrooms nonetheless presents teachers, curriculum developers and departments with the challenge of differentiating delivery of teaching material so that the right content is delivered in an appropriate way to the right children. Designing a curriculum that can cater for the needs of such a wide range of levels of achievement is extremely difficult. At the same time, these data beg the question how, given that all children are exposed to the same curriculum from age 5, is it that children at age 14 show such wide variation in their level of achievement in mathematics?

The scale of variation apparent amongst children in mathematics classrooms suggests both quantitative and qualitative sources of difference. There are clearly differences in the speed with which children acquire new concepts in mathematics. Also, as will be discussed below, there is increasing evidence of differences in the processes by which children develop new conceptions in mathematics. Aside from that evidence, though, it seems extremely unlikely that speed of concept acquisition alone is sufficient to account for the extreme variation in achievement level observed at age 14.

Given that teachers are attempting to scaffold their students' learning with appropriate interventions and support, it is imperative that teachers have access to and understanding of the ways in which children's learning processes can vary. The more information available regarding the ways in which children's mathematical development can differ, the more likely it will be that students' needs can be met in the classroom. If there are qualitative differences in the ways that children develop an understanding of mathematical concepts then there will be differences in the kinds of intervention and support required by those children in order to maximise mathematical development. If children with differing needs are taught mathematics as a homogeneous group, then there will necessarily be children who are disadvantaged, as they will not be receiving the teaching that they require in order to learn as efficiently as possible. Advances in technology are constantly allowing for greater and greater personalisation of learning. Research conducted now that can identify those variables that help describe as accurately as possible the differences in the ways that children learn, will take on some importance in informing the development of personalised learning systems.

The investigation reported in this thesis will go some way towards furthering an understanding of the ways in which children can differ in their development of mathematical concepts and potential sources of these differences. This will increase our understanding of the way in which children's learning of mathematics ought to be modelled. At some stage in the future it is envisaged that increasing individuality of learning will allow for children's personal needs for intervention and support in mathematics learning to be addressed. This investigation will help to define some of the variables across which children can differ in their learning of mathematics so that their individual needs can be identified as accurately as possible.

2.2 Strategy Variability

Students' strategy development in mathematics is an unquestionably valuable seam of research. Here, strategy is defined as the conscious logical process by which children generate a response to a mathematical problem. Mathematical problems often, if not always, have a range of strategies available for their solution.

"Strategy development" describes the process by which children make a transition from one use of strategy in response to a category of problem situations to a second use of strategy for the same category of problems. There is sufficient research in this area to suggest that interesting things happen during the transition between strategies. This section will describe some of the research that has investigated strategy development and show that a transition between different uses of strategy to solve problems is not a simple matter and that discussion of strategies in mathematics learning necessarily involves variability.

The traditional view of cognitive development is to view a child's mathematical development as a succession of increasingly more sophisticated conceptions. This can be thought of as a succession of stages, as in Piaget and Inhelder (1969), or as a staircase, as in Case (1992), where each successive stage or step represents a more advanced state of conception. These accounts of development have largely been validated by data collected at a macro level, involving many children and many trials. A major criticism of such models is that there is an overemphasis on the static states of children's understanding, correlated with age ranges, and that there is too little regard paid to the transitions between those static states.

Piaget describes cognitive change as resulting from disequilibrium, which occurs when a child's conception of the world does not fit with their experience. Disequilibrium is the catalyst for a restructuring of the child's conception of the world so that understanding is more consistent with experience. This explains *why* transitions occur, buts falls short of explaining *how* the transition between static states of knowledge occurs or what those transitional states look like.

Current research in cognitive development has largely rejected stage theories of development, in large part as a result of an accumulation of evidence of high levels of variability in children's thinking. Robert Siegler has conducted much of the most influential work in this area. His 'microgenetic' method (Siegler & Crowley, 1991), involving trial-by-trial analysis of strategy use by individual participants, has made it possible to show that children and adults are in transitional states a great deal more often then previous models of development have suggested. Children show a high level of variability within the individual, showing variability of strategy use for the same problem presented on consecutive days (Siegler & Shrager, 1984). Children can even show variability of strategy use within a single trial. Alibali and Goldin-Meadow (1993) showed that when children are solving mathematical equivalence problems (e.g. 5 + 3 + 4 = ? + 4), they often show different uses of strategy in their verbal explanations of solutions than they do through their gestures. Adults have also been shown to exhibit variability in strategy use in response to a wide variety of tasks including estimation (Dowker, Flood, Griffiths, Harriss, & Hook, 1996) and mental arithmetic (LeFevre, Sadesky, & Bisanz, 1996).

The three types of variability described above – between sessions, within a single session and within a single trial – mean potentially quite different things in terms of children's mathematical development. This is potentially problematic, given that each type of variability could equally well be described with the same phrase – 'strategy variability'. The research presented in this thesis will focus on variability between sessions. This is due to the focus on development in conceptual understanding; it is considered that questions of development can only be addressed between sessions, rather than within a single session. In this thesis then, 'strategy variability' should be taken to mean variability of strategy use between sessions.

Demonstrations, such as those described above, of pervasive variability in strategy use contradict the idea that there are stable, static states in children's conception of the world at all. The implication of such findings is that cognitive change involves not a quantitative shift in understanding, but a shift in the distribution of frequencies of strategies. An example from outside of mathematical development can help explain the difference between the Piagetian picture and the modern picture of cognitive change. Traditionally, in language development, children were thought to pass through a stage in which they over-generalised some past tense verb forms, using “goed” for “went, or “eated” for “ate”, for example. However, Kuczaj (1977) showed that between the ages of 2½ and 5, children produce both incorrect and correct forms of past tense verbs and rather than a sharp, qualitative change in behaviour, a gradual shift in frequency distribution from more incorrect to more correct usage can be observed.

The above findings have led to the development of the ‘overlapping waves’ model of development (Siegler, 1996) in place of variations on the stage model, which allows representation of the use of a variety of strategies at any one time, with the frequency of each strategy changing constantly. Within the model, new strategies can be discovered and those that are no longer useful can be removed from the repertoire.

The overlapping waves model also provides a framework to ask a number of new and interesting questions. Of central importance to this thesis will be the question of what individual differences are there in strategy variability, and what role might these differences play in the conceptual development of individuals? There is evidence to suggest that variability plays a key role in cognitive development. This evidence partly takes the form of studies in which periods of high variability are shown to be associated with conceptual change. Siegler (1995) shows in an investigation of number conservation that conceptual change is associated with an expansion and then contraction of a strategy repertoire. Van der Maas and Molenaar (1992), also investigating number conservation, show that a high level of strategy variability was associated with the transition between not conserving to conserving number. Church and Goldin-Meadow (1986) found that children who showed different uses of strategy in speech and in gesture were more receptive to teaching and, by implication more ready to learn, than children whose speech matched their gestures.

Further evidence for the importance of strategy variability for cognitive change can be taken from its central role in a number of theoretical accounts of development. A variety of models of change require that there are both mechanisms that produce variability in strategy use and mechanisms that produce adaptive choices between those strategies. For example, connectionist models involve the constant change of connection strengths between processing units from initial, random, levels, to final levels determined by the results of usage. Dynamic systems models also emphasise the importance of variability; “A dynamic approach elevates variability, both within and between individuals, into an essential element in the developmental process” (Thelen & Smith, 1994, p.341). In dynamic systems models, “variability is considered to be the harbinger of change” and “the essential ground for exploration and selection” (van Geert & van Dijk, 2002, p.342) In Piagetian models, change is caused by cognitive conflict – where children entertain competing strategies for the solution of a problem (Piaget, 1977).

With the accumulation of evidence for the importance of strategy variability in either causing or predicting cognitive change, it seems that the investigation of individual differences in strategy variability may be a good way to further an understanding of the reasons for differences in children's development of mathematical concepts. In fact, the investigation of individual differences in strategy variability should be expected to eventually shed some light on questions regarding the nature of the relationship between fluctuations in variability and cognitive change.

2.3 Individual Differences in Strategy Variability

If strategy variability is associated with cognitive change, then some of the variance in children's achievement in mathematics should be expected to be accounted for by differences in strategy variability. Findings regarding individual differences in strategy variability will help to describe differences in the ways that children learn mathematics, and help to highlight the differences between the ways that more and less successful students of mathematics develop. For this reason, the investigation of individual differences in strategy variability must be considered worthwhile.

Arguably, a more important factor in strategy use than the number of strategies available is the way that problem-solvers choose between those available strategies. One reason for this is that it would seem unlikely that a successful strategy would be *forgotten* at a later date. It is more likely that the strategy remains in the repertoire but is no longer called upon for solving problems due to the knowledge of alternative, more efficient strategies. It can be argued that when strategy choice is investigated experimentally, this 'strategy adaptivity' better describes observations than strategy variability.

Torbeyns, Verschaffel and Ghesquiere (2006) show that high achievers in mathematics are better able to choose strategies according to problem demands and strategy performance characteristics than other students. This is an indication that individual differences in strategy variability are likely to be interesting both for the researcher and the classroom practitioner. For example, Torbeyns et al. discuss the possibility of increasing adaptivity in the classroom through conceptual and investigative instructional approaches.

The reason that the term ‘strategy variability’ is used in this thesis rather than ‘strategy adaptivity’ is that ‘adaptivity’ implies that children are making a choice amongst a number of strategies on the basis of some criteria, possibly efficiency or the probability of arriving at a correct answer. In the research described in this thesis, there is no analysis of the efficiency or comparative success rates of strategies. The analysis will be concerned only with the range and distribution of strategies employed by children in order to solve problems. Therefore ‘variability’ – in this case – is a more suitable term than ‘adaptivity’.

Schunn, Reder, Nhouyvanisvong, Richards and Stroffolino (1997) investigated participants’ use of retrieval and computation strategies for solving arithmetic problems. They found that all participants used both strategies, and that strategy use was adaptive. That is to say that as participants became more familiar with problem stimuli, they were more likely to use a retrieval strategy. As might be expected, more unfamiliar problems were likely to be solved using a computation strategy. The interesting thing about Schunn et al. (1997) is that the methods they used allowed them to conclude that the decision between methods was made before either solution process began.

Schunn and Reder (1998) built on Schunn et al. (1997) with an investigation of individual differences in strategy adaptivity, showing that adaptivity in a variety of domains varies more across individuals than should be expected as a result of sampling noise. This is one way to demonstrate that individual differences in strategy variability exist. However, it is limited in the sense that it does not tell us anything about why those differences exist. One intended end result of mathematics education research is that in the classroom, teaching can be designed so as to ensure that students can each be given the kinds of support and intervention that as closely as possible meets their needs. If children have differing styles or levels of strategy variability, then it is likely that different kinds of support and intervention will be required for different groups of children so that each child is best able to reach their highest potential level of achievement in mathematics. The more that we are able to find out about the way that levels of strategy variability are distributed across children, the more able we are to match delivery of teaching to the needs of children.

This thesis aims to take a different approach to the identification of individual differences. The process under investigation is the development of strategy use over time. More specifically, this investigation will focus on the differences between students who choose to use a similar set of strategies and those who use a different set of strategies in response to a problem set presented for the second time than that used on the first occasion.

Now, the expectation is that some participants will use a similar set of strategies and some will use a different set of strategies in a second session than the set used in the first session. A major aim of this thesis is to identify factors that predict these patterns of strategy use over time. Therefore, within this thesis, individual differences are conceived of as the sum of group differences. There are a number of group factors that give any individual their individuality – these might include sex/gender, race, socio-economic status, education level and background, personality type, working memory capacity, maths-fact retrieval facility and processing speed for example. Any number of these group factors may have an impact on a particular behaviour that we are able to measure – in this case pattern of strategy use over time. If there are individual differences in patterns of strategy use, then these are made clearer and more useful with the identification of contributing group factors.

2.4 Potential Group Factors Influencing Strategy Variability

2.4.1 Gender

In the specific case of strategy variability in mathematics, some group factors can be highlighted as likely sources of variation in patterns of strategy use. Firstly, we can consider those factors where there is a strong evidence base for an effect on strategy use. Although there has been very limited research on factors affecting strategy use over time, there does exist a body of work on strategy choices for individual problem sets. One factor that stands out in predicting differences in strategy use is gender. Several studies (e.g. Carr &

Davis, 2001; Carr & Jessup, 1997; Carr, Jessup, & Fuller, 1999; Davis & Carr, 2001; Fennema, Carpenter, Jacobs, Franke, & Levi, 1998; A. M. Gallagher & de Lisi, 1994; A. M. Gallagher et al., 2000) have shown that there are differences in the ways that boys and girls select strategies for solving a variety of mathematical problems. This suggests that it will be worthwhile investigating the effect of gender on strategy variability.

Of course, there is extensive evidence in the literature regarding performance differences according to gender (e.g. Fan, Chen, & Matsumoto, 1997; A. Gallagher, 1998; Geary, 1996; Halpern, 1986; Hyde, Fennema, & Lamon, 1990; Johnson, 1996; Kimura, 1999; Maccoby & Jacklin, 1974; Mullis, Martin, Fierros, Goldberg, & Stemler, 2000; Reis & Park, 2001; Sammons, 1995). The literature regarding gender differences in mathematics will be discussed more thoroughly in section 2.5.

2.4.2 Race-ethnicity and Socio-economic status

In addition to gender, ethnicity and socio-economic status have some evidence in the literature to suggest a role in predicting performance in mathematics. Secada (1992) analysed the mathematics achievement of a range of groups, categorised according to ethnicity, social class and language proficiency. The primary test used in order to measure achievement in mathematics was the NAEP (a basic-skills examination administered to a sample of students at ages 9, 13 and 17 across the US periodically since 1973). Findings indicated that for each of the three age groups, White children perform better than do African-American or Hispanic children. There was some evidence that the achievement gap was closing, but Secada (1992) considered that this might be partly due to the fact that basic skills (which make up the abilities tested in the NAEP) are the skills that are highlighted in ‘catch-up’ programmes designed to support students, especially in minority groups, struggling with mathematics.

In analysing the relationship between SES and mathematics achievement, Secada (1992) considered measures including parents’ level of education and the community in which a student lives. On both measures, students from higher SES groups were shown to outperform students from lower SES groups by a substantial margin.

Tate (1997) again analysed score on the NAEP assessment in order to provide an update on Secada (1992). Findings indicated that between 1973 and 1992, White, African-American and Hispanic students all showed an increase in performance. White students performed at the highest level of the three groups, but between 1973 and 1992, the gaps between groups had narrowed somewhat. Using the same assessment scores, Tate demonstrated a strong relationship between SES and mathematics achievement. In addition, there was an interaction between ethnicity and SES whereby low-SES students from minority ethnic groups showed particularly low levels of mathematics achievement.

There seems to be little or no literature available regarding differences in use of strategy according to ethnicity or SES. The closest approximation to this is probably from the situated learning literature, such as Carraher, Carraher and

Schliemann (1985), in which studies show that children's ability to solve problems depends on the similarity of the context of a given problem to the context in which children learned the relevant concepts. Children from different ethnic groups and from different social classes are exposed to different environments for mathematical learning, and therefore are likely to be familiar with different contexts of mathematical problems. However, this association is still more about performance than strategy use.

2.4.3 Summary

The exploration of individual differences in variability in the first study will focus on gender differences in strategy variability, as this is the factor with the greatest weight of evidence for a relationship with differences in performance and in strategy use. Other group factors that could be considered for testing, such as ethnicity and SES do not have the same weight of evidence for independent, pervasive causation of differences in performance, nor do they have a backing in the literature to suggest differences in strategy use.

If it is possible to show that gender can account for some of the individual differences found in strategy variability, that will be a big step forward in understanding the reasons for differences amongst children in mathematical development. Gender is a major factor in educational research and is present as a discriminating factor between students in most mathematics classrooms. Any findings regarding an association between gender and learning processes will likely be a significant addition to the literature.

The purpose of the investigation to be reported in this thesis is to try to explain some of the individual differences that have been identified in strategy variability previously. It seems that gender offers a potential explanation for a significant portion of individual differences in variability. Therefore the next section involves a more thorough analysis of the literature relevant to discussions of gender differences in Mathematics.

2.5 Gender Differences in Mathematics

There is a fairly long history of research regarding gender and mathematics education, which has shifted in focus somewhat over the last decade from the general to the more specific.

2.5.1 Achievement

Maccoby and Jacklin (1974) claim that after the age of 12, boys exceed girls in mathematical ability. Halpern (1986) claims that amongst children aged between 13 and 16, boys consistently outperform girls in tests of mathematical ability. However, more recent research has shown that such sweeping generalisations are not justified. While researchers continue to find gender differences in mathematical achievement, the effects of gender are more limited than suggested by some early work.

Hyde, Fennema and Lamon (1990) conducted a meta-analysis of gender difference research, involving analysis of 259 studies, and generated a number of interesting findings. While they calculated an average effect size of 0.20 favouring males, they found that the differences between genders depends to a large extent on at least three other factors; including the cognitive level of the test, the selectivity of the sample, and the age of the students. Gender differences were most apparent for tests involving problem solving – in fact tests involving computation tended to favour girls. Differences were also most apparent for more selective samples. The greater the extent to which samples were selected on the basis of ability, the larger the difference in achievement between boys and girls. Finally, differences were only apparent in studies where children were aged 15 or above (this is one finding that is reflected in the claims of both Maccoby and Jacklin (1974) and Halpern (1986)). Hyde, Fennema and Lamon (1990) also note that there has been a decline in gender differences in mathematics since 1973.

Hedges and Nowell (1995) found some similar results in terms of the selectivity of samples, in that they found that while there were negligible differences between boys and girls amongst the general population, amongst the top 10% of achievers, boys outnumbered girls. They noted that while gender differences amongst the general population have been decreasing over time, the ratio of boys to girls amongst the highest achieving students has remained relatively stable. Hedges and Nowell argue that the greater proportion of boys than girls amongst the top 10% of achievers is due to a combination of a slightly higher mean and a higher variance in ability.

The TIMSS data, collected in 1994-5, has revealed widespread gender differences in mathematics achievement (Mullis et al., 2000). Some small differences in overall achievement between boys and girls were found amongst 9-year-olds and 13-year-olds. For those countries in which there was a significant difference, boys achieved a higher score than did girls. Analysis within content areas shows that gender differences are particularly visible for questions involving measurement and estimation.

The TIMSS data show an increasing gender difference in overall achievement in mathematics with age. Amongst students in the final year of secondary education, scores for both ‘mathematical literacy’ (for all students) and ‘advanced mathematics achievement’ (students taking advanced mathematics courses) were significantly higher for boys than girls in the majority of countries surveyed. These findings, regarding increasing differences between genders with age, correspond closely with those of Hyde, Fennema and Lamon (1990) discussed above.

What conclusions can be made on the basis of the literature described so far? It seems that there certainly are some differences in levels of achievement between boys and girls in mathematics. It is important though to realise the bounds of these differences. Differences in achievement have only been shown to exist within certain content areas, for certain problem types and only once students reach a certain age. In addition, reported effect sizes are always such that between-groups variability is dwarfed by within-group variability. Nevertheless, even with such limits imposed on the conclusion, it is clear that further investigation of gender differences is warranted.

2.5.2 Participation

There is a clear difference in the numbers of boys and girls choosing to continue taking mathematics courses in post-compulsory education in the UK.

For example in the year 2003/4, the most recent year for which confirmed figures are available, 32,078 boys entered examinations in A-level mathematics compared with 19,050 girls (DFES, 2005a). This is despite girls' entries for all A-level subjects outnumbering boys' 363,673 to 312,254. For boys, mathematics A-level is the most popular course, while for girls mathematics ranks 8th. The most popular A-level subject for girls is English Literature.

It is difficult to interpret these figures as it is not possible to know whether lower participation is due to lack of ability (real or perceived), lack of enjoyment, consideration of future career or any other factor. They do suggest however, that there are real differences requiring further investigation between boys' and girls' relationships with mathematics.

2.5.3 Cognitive and Affective Factors Associated With Gender Differences in Mathematics

During the last decade, research has shifted in focus to some extent. While some studies continue to investigate differences in achievement and participation in mathematics at various levels, there has been an increasing amount of research investigating various factors that might be responsible for those differences. This section will describe some factors that have been identified as having potential to explain some of the differences in performance in mathematics between genders. These include mental rotation, math-fact retrieval, anxiety and motivation.

A large amount of research has been conducted investigating mental rotation as a mediator of gender differences in mathematical ability. Research has shown both that there is a difference in mental rotation facility between boys and girls (Masters & Snaders, 1993; Voyer, Voyer, & Bryden, 1995) and that facility with mental rotation is a mediator of gender differences in ability in some mathematical domains (Delgado & Prieto, 2004), including geometry and word problems. There may be a difference in the likelihood of boys and girls using spatial strategies for problems of a spatial nature (Delgado & Prieto, 1997) – suggesting that there may be differences between boys and girls in the application of strategies to the kind of graph problems used in the studies reported in this thesis. Delgado and Prieto (2004) also investigated the mediating effect of lexical access on the gender effect on mathematics achievement and found that facility with retrieval from the lexicon is a mediating factor in problems involving arithmetic, geometry and word problems.

Math-fact retrieval has also been identified as a potential source of differences in problem solving ability. Children with good facility in math-fact retrieval are able to quickly recall the answers to simple arithmetic problems. This is thought to affect performance in mathematics in two ways. Firstly, any standardised mathematics test will involve a number of instances where arithmetic facts are required. The more quickly that these can be brought to mind, the more time remains to work on the remainder of problems in the test. Secondly, during solution of problems, cognitive load is reduced by automatic processing of low level arithmetic operations. Greater facility with math-fact retrieval creates greater remaining cognitive capacity to use for higher level problem solving processes. Royer, Tronsky, Chan, Jackson and Marchant

(1999) conducted a series of studies in which they were able to show that facility with math-fact retrieval is both a good predictor of performance on problem solving tasks and is itself predicted by gender. They concluded that boys' greater facility in math-fact retrieval accounted for a significant portion of the variance in children's level of achievement in mathematics tests.

The math-fact retrieval hypothesis has been criticised (Geary, 1999; Wigfield & Byrnes, 1999) for failing to explain differences in performance on problems that do not involve the use of arithmetic facts, such as geometry. Geary (1999) argues that math-fact retrieval cannot be the primary source of the male advantage in mathematics performance in the way that mental rotation can, for this reason.

Math anxiety is an example of an affective factor influencing performance in mathematics. Arguments concerning the effect of anxiety on performance are structured in much the same way as one of the arguments above, concerning math-fact retrieval. Anxiety specific to mathematics has been shown to affect performance by reducing the working memory capacity available to work on problems. Ashcraft and Kirk (2001) showed that anxiety affects mathematics performance by reducing working memory capacity. Miller and Bichsel (2004) showed anxiety to be a strong predictor of performance and that this was due to the effect of math anxiety on visual working memory performance.

Math anxiety has also been shown to be more common amongst females than males (Hembree, 1990; Lussier, 1996; Miller & Bichsel, 2004). Therefore it appears that math anxiety is likely to be a strong mediator of gender differences in performance in mathematics.

2.5.4 Gender Differences in Strategy Use

There is a growing literature on gender difference in strategy use in mathematics. Much of the work done up to now has involved children in their first years of education. For example, Fennema, Carpenter, Jacobs, Franke & Levi (1998) conducted a longitudinal study, following children through the first three years at primary school and asking them to solve simple addition and subtraction problems. They found no differences in proficiency between boys and girls, but they did find some strong differences in the strategies used to solve the problems. Girls tended to use more concrete strategies, involving modelling or counting, while boys tended to use more abstract strategies, such as invented algorithms. By the third year of primary school, girls were much more likely to use standard algorithms for solving addition and subtraction problems than were boys. For both boys and girls, those children that used invented strategies for solving problems were better able to solve extension problems than those who did not.

Carr & Jessup (1997) also found differences in the strategies used by first-year primary school boys and girls to solve simple addition and subtraction problems. They found that girls were more likely to use overt algorithmic strategies such as counting on fingers while boys were more likely to use retrieval, answering from memory. Significantly, boys developed a preference for retrieval before they were able to use this strategy to generate correct answers. This corresponds with Siegler's (1988) finding that girls tend to be perfectionists, using strategies that are guaranteed to produce correct answers, at the risk of inefficiency. This also corresponds well with Royer et al. (1999) mentioned above, who found boys to have greater facility in retrieving arithmetic facts from a young age,

When strategy use was controlled for so that all children used retrieval, boys outperformed girls, indicating that strategy choice is associated with both skill and preference (Carr & Davis, 2001). This is important when considered alongside emerging evidence that when children can easily retrieve arithmetic facts they are also more likely to possess conceptual knowledge about mathematics (Canobi, Reeve, & Pattison, 1998).

Davis and Carr (2001) found that boys' and girls' strategy choices were differentially predicted by temperament, whereby boys' strategy use was influenced by levels of impulsivity and girls' strategy use was influenced by levels of inhibition.

Although the above studies involved participants at a much younger age than those in the present work, and the problems were simpler, it seems that there is a growing body of evidence to suggest that boys and girls do mathematics differently. It also seems that boys and girls do mathematics differently for a variety of reasons. These findings must be borne in mind when considering the implications of findings reported in this thesis.

There has been a small amount of work done investigating gender differences in strategy use amongst older children. Gallagher and de Lisi (1994) conducted a study in which they interviewed students who had recently taken the SAT-math paper – a university entrance examination taken by high school students in the US. Gallagher and de Lisi categorised problems in the SAT-math paper as either conventional, requiring clearly defined algorithmic methods, or unconventional, requiring an atypical solution strategy such as an unfamiliar use of an algorithm or some type of estimation or insight. They found that male students were more likely to use an unconventional strategy to solve an unconventional problem than were girls. A replication of this study was conducted (A. M. Gallagher et al., 2000) which led the researchers to conclude that strategy flexibility, the ability to draw on a variety of strategies to match problem demands, is a source of gender difference in mathematical ability.

In the context of variability, the literature discussed in this section suggests that boys should be expected to show higher variability across sessions than girls. Both Fennema et al. (1998) and Gallagher et al. (2000) suggest that girls are more likely to use algorithmic, rote-learned strategies in response to problems than are boys. While participants in the research reported in this thesis will not have been introduced to the experimental problem set in the classroom, it is likely that the tendency observed in these studies, of girls to repeat the use of strategies across sessions, will be observed again here.

2.5.5 The Gender Perspective Within This Thesis

It is clear from the research described in this section that some gender differences exist. As discussed above, however, it is important to be sceptical when discussing gender differences in mathematics for a number of reasons, not least the fact that any between-groups differences are often small in comparison with within-groups differences. The literature discussed above is sufficient, however, to suggest both that gender is a likely source of differences in variability of strategy use and that further analysis will be of interest to the research community.

In line with the perspective taken within this thesis regarding strategy use and conceptual understanding, the perspective on gender will be that attention will be applied to what children do rather than what they do not do. In other words, boys and girls will not be measured against one another. As Gilligan (1982) argues, it is not appropriate to use either gender's (usually male) performance as a benchmark against which to measure the performance of another (usually female). In this thesis, gender will be used as means of opening up questions of individual difference in strategy development, not as a means of producing generalisations.

This then is not the place to discuss equity issues, except to reinforce to position held throughout the thesis, that the purpose of mathematics education should be an attempt to maximise learning of mathematics for every student. The way to do this is by assessing and meeting the learning needs of each individual. The study of gender differences in mathematics can help to identify items within the realm of mathematical understanding in which individual differences can be found.

Finally, discussion of the origins of any gender differences in mathematics performance may seem conspicuous in their absence form this section. There are two reasons for this absence. Firstly, arguments regarding possible biological and social explanation for gender differences in cognition are as common as firm conclusions on the subject are few. Secondly, the origin of gender differences, while interesting, is not thought to have a huge bearing on the questions raised in this thesis. For the remainder of this thesis, it will be sufficient that an understanding of what differences exist in problem solving behaviour between boys and girls is established. Discussions about the origins of those differences are likely to be supplementary and likely to complicate rather than complement the key issues in this thesis. The role of gender in this thesis is to act as a probe for exploring individual differnces in strategy variability. To that extent, other variables could have been investigated, such as race or SES (discussed above), IQ or mathematical ability. The reason gender has been chosen for investigation is that there is a growing literature suggesting gender differences in strategy choice.

The aim of this research is not to argue for differentiated teaching according to gender, but to provide further evidence for individual differences in strategy variability. If a gender difference in variability is found, it will provide a foundation for further exploration of these individual differences in mathematical development.

2.6 Rate of Change Problems

Rate of change problems have been selected for this investigation for a number of reasons. One significant reason is that they are the type of problem used in Mevarech and Stern (1997). The investigation reported in this thesis begins with an augmented replication of one of the experiments reported in Mevarech and Stern and subsequent studies were designed to elaborate on the findings of that replication. However, rate of change problems have characteristics that make them suitable for the investigation of strategy use. As will be shown below, they are problems that children have access to, in that they are able to give a meaningful answer, but have difficulty answering correctly. Children have a wide range of available strategies for answering these problems. Some of these strategies, if applied to the problem correctly, will give a correct answer. Some strategies will sometimes give a correct answer and sometimes give an incorrect answer, depending on certain aspects of the problem. The remainder of strategies used will generally give an incorrect answer. Examples will be given later, in section 3.1.1. The wide range of available strategies will mean a greater likelihood of exposing differences in patterns of strategy use amongst participants.

Rate of change problems as discussed in this thesis are a variant of a more general category of line graph problems. The solution of even simple graph problems seems from the literature to be a complex task, subject to numerous errors.

Current models of graph comprehension involve three integrated processes. These include pattern recognition processes, interpretive processes that construct meaning from those encoded graphic patterns, and integrative processes that associate constructed meanings with referents derived from titles and labels (Carpenter & Shah, 1998). Carpenter and Shah used eye-tracking software to show that the interpretation of graphs involves cycles through these three processes, encoding patterns to reveal x-y functions then identifying variables associated with them. This model implies a high level of variability in children's strategies for solving graph problems, as there are several occasions on which children's problem solving processes can diverge. Children might identify different graphical patterns from one another, they might interpret them in different ways, and they might associate these with titles and labels in different ways.

The place of these kinds of problems in the UK National Curriculum is important, as it will help to determine a suitable sample for the investigation. Understanding rate of change problems involves aspects of each attainment strand in the curriculum, 'number and algebra', 'shape, space and measures' and 'handling data'. Table 2 lists excerpts from the National Curriculum attainment targets that pertain to rate of change problems, by level of attainment.

Table 2: Attainment targets in the UK National Curriculum pertaining to rate of change problems

Attainment Level	Attainment Strand	Excerpts from National Curriculum
4	Number and Algebra	Pupils use and interpret coordinates in the first quadrant
	Handling Data	They construct and interpret simple line graphs
5	Number and Algebra	Pupils use and interpret coordinates in all four quadrants
	Handling Data	They interpret graphs and diagrams and draw conclusions
7	Shape, Space and Measures	Pupils understand compound measures, such as speed

The children that this investigation will be concerned with are those who have access to the problem sets, but do not have a well-formed concept for their solution. The information above suggests that children at level 5 will have a good understanding of the requisite prior knowledge for the problem sets, but will not have a good concept of the relationship between the shape of a linear graph and the rate of change in quantity represented.

Curcio (1987) found no gender difference in children's understanding of mathematical relationships expressed as graphs. This is important, as in this investigation the focus is on use of strategy rather than performance. It is useful that the domain under investigation has not in the past shown gender differences in achievement, as this will support arguments later that differences in behaviour are independent of students' ability to understand the problem sets.

2.7 Theoretical Framework

The reason for including this section is a desire to make clear the context in which this research is being conducted. Interdisciplinary research can run the risk of being misunderstood if this step is not taken. This aim of this section is to make clear what perspective on research on mathematics education is being taken here, and why.

The first part (2.5.1) compares two very different perspectives on research in mathematics education and discusses a way in which the two perspectives can be combined and resolved. The remainder of the section details methodological issues arising in this investigation, including sampling and ethical considerations.

2.7.1 Situationism and Cognitivism

Are situationism and cognitivism compatible or are they incompatible? If they are compatible, then how should we describe the conjunction? If they are not compatible then which is the best? At the lowest level, these seem to be the questions that we are trying to answer. The first of the three questions looks like a good place to start.

Maybe a good way to start might be to have a look at some of the principles of each of the two approaches and see if any of them suggest incompatibility. In order to do this though, we are already making the assumption that both approaches have the same frame of references, that both deal with the same sorts of things. Maybe it would be a good idea to establish the frame of reference we are interested in first then.

What we want from any theoretical approach is a framework for understanding the ways in which children develop mathematical understanding. We want to be able to take guidance from some approach in order to aid the formation of research questions, the making of decisions about methodology, and the inference of conclusions from analysis. It seems that both situationism and cognitivism have something to say about both of these areas. It should be useful to consider what both approaches say about the kinds of questions we should be asking, the kinds of research we should be doing and the kinds of inferences we should be making.

The picture that situationism gives us is that children exist in a community of practice in which they are learning mathematics by participating as ‘apprentice’ mathematicians or by participating in a classroom community that is developing its own mathematical practice (e.g. Lave & Wenger, 1991). Children’s learning of mathematics is a construct of their increasing participation in mathematical practice. The focus of situationism is very much on the observable behaviour of children within a community of practice. Therefore the kinds of research questions that we need to ask are those that compare changes in situation with changes in behaviour. In the classroom we may be interested in the kinds of teaching that result in desirable behaviours, or we may be interested in behaviours that are reproduced in more or less similar situations - of course, the variables that make up a situation will be numerous.

When taking a cognitive approach, we are instead asking questions about the kinds of processes that occur in children's minds as a result of problem situations. We want to know what kind of processes happen in response to what kind of problems, and how similar or different those processes are when a problem situation changes.

The perspectives of the two approaches seem to be very different. For the cognitivist, meaning in mathematics is considered to be created by the individual. For the situationist, meaning first exists as a social object, to be internalised by the individual through participation in mathematical practice. This central difference in perspective leads to further differences in theory and practice. For the cognitivist learning takes place as a result of physical interaction with the world, while for the situationist, learning takes place as a result of social interaction with a community. For the cognitivist, teaching through abstraction can effect learning. For the situationist, learning is effective only in 'authentic' situations, in complex social environments.

It seems that there are three choices. We can take the cognitive approach, we can take the situational approach, or we can make some compromise between the two. Anderson, Reder and Simon (1996; 1997) and Greeno (1997) seem to take the position that it is right to take one or other approach – Anderson et al. argue for the cognitive approach, while Greeno argues for the situative approach.

Lerman (1996; 2000) and Steffe and Thompson (2000) argue for and against the compatibility of the two approaches. Lerman believes that it is not possible to integrate a sociocultural view of learning with a constructivist theory of learning. He claims that social cognitivist and social constructivist theories are incoherent attempts to incorporate intersubjectivity into constructivism, "I suggest that the extension of radical constructivism toward a social constructivism, in an attempt to incorporate intersubjectivity, leads to an incoherent theory of learning" (Lerman, 1996 p.133).

I think that I would tend to agree that it is not possible to agree with both points of view at the same time. Meaning in mathematics cannot derive exclusively from the individual and at the same time exclusively from the sociocultural plane. I also think that it is awkward to try and manufacture some kind of compromise between the two. It seems to me that attempts to integrate the two can only be successful if one or both is weakened to a fairly high degree.

What I do think is possible is that both approaches can be maintained as they are and can both be applied to research situations. Both can be part of the research toolkit and used as appropriate. I also think that this can be done without contradiction and without incoherence. When looking at some particular teaching and learning situation it should be beneficial to look from both the cognitivist and the situationist perspective. It probably doesn't matter which of the perspectives is the 'right' one, it might not even make sense to ask the question. It is the case, though, that each has the potential to aid the researcher in making decisions.

To give an example, I will focus on the question of transfer. The cognitivist frames this question in terms of the knowledge of the individual, and how well that knowledge can be utilised in a variety of problem situations other than that in which it was acquired. This approach leads us to ask questions about how knowledge is constructed by the individual so that it can be recalled and adapted to suit new problem situations, e.g. by analogical reasoning.

The situationist frames the same question, by asking what kinds of situation are likely to bring about the same kinds of behaviour as the target situation. Considering the investigation to be reported in this thesis, it seems that each perspective will be required at different stages. In framing a research question, a situationist perspective has been taken, in that a situational variable (gender) and its effects on children's response to problems has been considered. In the design for the experiments, a cognitivist approach will be taken, using children's explanations of their answers in order to infer strategies used.

How can the use of both perspectives without dilution be justified, however, when the incompatibility between them has been so often discussed? It might be useful to take a step back. The question we are trying to answer is 'what is learning?' possibly more usefully phrased as 'what happens during learning?'

Clearly there are things that happen in the mind of a student during the development of understanding; knowledge and understanding change with experience of problem situations. Investigation and analysis of such changes in knowledge and understanding depending on experience must be useful in informing teaching.

The arguments of Lerman and Greeno stem from an idea that whatever happens inside the mind of a student is outweighed by events that take place outside the mind, in the community to which the student belongs, and that changes in a student's knowledge and understanding are a necessary result of events that take place in that community. This is, as others have commented, a behaviourist picture of development. There is also a misinterpretation of constructivism involved, leading critics to argue that constructivists ignore social aspects of learning. In fact constructivists maintain that social interaction is a crucial source of learning. For example von Glasersfeld (1995) states that social interaction is both the most common cause of cognitive conflict and the most useful way of testing constructions.

It is almost certainly the case that neither perspective can ever provide a complete picture of development independently of the other. Each has the potential to describe a system of some kind but that system requires complex input from the neglected source (the individual mind in the case of situationism and the community in the case of cognitivism). Therefore the system described from either perspective is neither independent nor complete. Each perspective is dependent on simplified input from its respective neglected counterpart. The cognitivist group uses problem situations in a laboratory-like environment in place of complex communities of practice while the situationist group uses automata in place of thinking beings. There is nothing wrong with either simplification. The point at which the individual interacts with the world is extremely complex – possibly to the point of being impossible to study directly.

It seems that if all this is true, then it is meaningless to argue for the predominance of one perspective over the other. Equally, it is meaningless to argue that one or other ought to be diluted in some way in order to achieve a compromise. The two perspectives are neither compatible nor incompatible, as there is no point at which both can be taken – each stops at the point of interaction. Any one specific question ought then to be answered within one or other perspective, dependant on the context of that question.

The perspective taken in this thesis will be that questions in mathematics education can be answered using each perspective individually, but in a limited way. For coherent and comprehensive picture of teaching and learning in mathematics, research must be open to both perspectives. Within this thesis the dominant perspective will be cognitivist, as the questions asked are generally to do with the individual student's thought processes. However, it is important that the investigation is written in such a way as not to prejudice the work against future developments from the situationist perspective. When discussing questions of potential implications of this research, or the potential for future research, efforts will be made to leave the door open for research from both perspectives.

2.8 Methodology

From previous research it is reasonable to make the claim that there are individual differences in children's strategy variability and that these will be accounted for by a number of yet to be identified group factors (see sections 2.2 and 2.3). This claim of the existence of one or more causal relationships means that the research paradigm to be used will be postpositivism. Postpositivist researchers hold that a single reality exists and the job of the researcher is to know that reality as well as possible. Research is conducted in the context of a predetermined theory, and the intent is to make a stronger case for that theory.

Within the postpositivist paradigm, objectivity is very important in order to ensure a focus on the relationship in question, unaffected by extraneous variables. Generally, postpositivist research is associated with quantitative methods. Historically, positivist researchers in psychology and education have used experimental methods, borrowed from the natural sciences. The experiment often seems to be the default methodological tool for the psychologist. Robson (1993) cites Fisher (1935) as a 'major influence' in the development of the experiment. The essence of an experiment is the comparison of random samples from known populations, enabling the use of statistical theory to test for significant differences. A problem with the use of experiments is that 'random sampling from known populations, although in principle still feasible, appears to present extremely difficult practical and ethical problems' (Robson 1993, p.46) when used in research in the real world. As a result of such problems, quasi-experimental methods have been developed

(Cook & Campbell, 1979) that accept the problems of non-random sampling and allocation to groups.

Experimental methods are not appropriate for the planned study. An experiment involves two or more groups treated in different ways (conditions of the independent variable) that are then measured according to some criteria (the dependent variable). The conditions of the independent variable are the hypothesised causes or predictors according to theory being tested, plus a control group where possible for the purposes of comparison. The dependent variable is then the hypothesised effect according to the theory.

It is not possible to treat groups of students in different ways according to the independent variable of the planned study. Any potential group factors that might be investigated will be pre-existing in students and therefore will not be manipulable by the researcher. This rules out the use of both true- and quasi-experimental methods. Positivist studies in this field rarely use experimental methods, and generally fall into two categories - either correlational or causal comparative studies. The main difference between these types of study is the type of conclusion that can be drawn. Causal comparative studies, like experiments, test for differences, while correlational studies test for relationships between variables.

The causal comparative method has been used in determining changes in mathematical cognition with age. Girelli, Lucangeli and Butterworth (2000) conducted a study investigating the development of automatic numerical processing between the ages of 6 and 8. Children aged 6- 8- and 10-years-old were given tests to elicit a version of the Stroop effect with numbers. In one test, the children were shown a card with two numbers on and asked to choose the physically bigger number, ignoring the numerical magnitudes of the numbers. 8-year-old children showed a greater effect of magnitude than 6-year-olds, and 10-year-olds showed a greater effect than 8-year-olds. The researchers argued that automatic processing of the numerical magnitude of the numbers was interfering with children's selection of the physically bigger number, and concluded that automatic processing of numerosities develops between the ages of 6 and 8 and continues to develop at least until the age of 10. In this study, neither variable – the independent variable being age and the dependent variable being the measure of the effect of magnitude – were manipulable by the researchers.

The correlational method has been used by researchers trying to find predictors for mathematical achievement such as processing speed or working memory utility. Gathercole, Pickering, Knight and Stegmann (2004) compared 7- and 14-year-old children's national curriculum assessments with scores from a battery of complex working memory span tasks. The researchers found that attainment levels in English, Maths and Science were positively correlated with working memory scores at age 7 and levels in Maths and Science were strongly correlated with working memory scores at age 14. They concluded that achievement across the curriculum is constrained by working memory capacity across the childhood years. So, in this study neither variable – national curriculum assessment nor working memory capacity – was manipulable by the researchers. The difference in this study was that Gathercole et al. found a relationship between two variables rather than a difference in one variable dependant on another.

The early stages of the investigation presented here consist, in a large part, of a replication of work by Mevarech and Stern who conducted an experiment investigating the transfer of reasoning skills between sparse and realistic contexts in a particular set of mathematical problems. They devised a graph problem, with three questions about rate of change, which could be presented either with or without supporting context (the profits of two businesses). One group of students was asked to solve the problem with sparse context then the problem with a realistic context. A second group solved the same problems, but in reverse order. They found that students who solved the realistic-context problem first attempted to use practical reasoning skills to find solutions, and then tried to use the same skills to solve the sparse-context problem (not possible), resulting in poor performance. Student solving the sparse-context problem first used logical and mathematical reasoning skills, which they transferred to the solving of the realistic-context problem, resulting in better scores (Mevarech & Stern, 1997). However, not all students participating in the study behaved in the same way. There were differences in the way that students transferred knowledge between the two contexts, and many students were able to provide correct solutions in both conditions. Assuming that the results will be replicated in the planned study, I will want to know how much of the variance in strategy variability can be accounted by for one or more yet to be identified group variables.

Initially, an exploratory study will be conducted, aiming to test for individual differences in strategy variability. If individual differences are found, analysis will be conducted involving the testing of potential group factors' association with levels of strategy variability.

Having identified one or more group factors associated with strategy variability, causal comparative studies will be used in order to establish the robustness of those relationships.

2.8.1 What's wrong with postpositivism?

While the methods of the postpositivist paradigm have much to offer, there are aspects of inquiry that are inaccessible using methods discussed up to now. One major aspect that is relatively inaccessible within the postpositivist paradigm is process. Experimental methods can give the researcher a good deal of confidence in the existence of a relationship between age and ability or between teaching methods and educational outcomes, but are lacking regarding questions about the processes by which learning is achieved. Siegler warns against making claims on the basis of experimental research and argues that taking averages, as in most positivist research, can produce misleading conclusions. In a study involving young primary school children's addition, Siegler found that if latencies were analysed together for any child, his data supported previous research claiming that children at this age consistently use the *min* strategy, counting on from the higher number. However, when verbal reports from children were analysed alongside solution times a different picture resulted. Children used up to five distinct strategies in solving addition problems – most used at least three. No one strategy was used on more than 40% of trials (Siegler, 1987).

Similar to the issues raised by Siegler are Gelman and Gallistel's criticism of Piaget's methodology (Gelman & Gallistel, 1978). Piaget set out the cognitivist explanation of the child's understanding of number on which much subsequent work has been based. The stages of development of a number of requirements necessary for a true understanding of number are defined on the basis of a series of clinical interviews. Those requirements of understanding consist of:

- Conservation of continuous and discrete quantities
- Correspondence/equivalence between sets – leading to an understanding of the cardinality of a set
- Additive composition of number
- Reversibility of operations on number

(Piaget, 1952)

Many subsequent studies regarding the development of children's understanding of number and of arithmetic have been essentially Piagetian in nature. For example, '…[constructivism] has evolved from the basic Piagetian notion that individual's *construct* their knowledge as they interact with the world' (Nickson, 2000). Constructivists advocate that the best way to teach mathematics is through problem solving, therefore giving children the opportunity to construct their own knowledge from experience. However, researchers are aware that the child does not construct his knowledge of the world independently, and some (e.g. Cobb, Wood, & Yackel, 1991) have used the term 'social constructivism' in acknowledging the fact that a great deal of success in the classroom depends on the relationships between teacher and students.

Gelman and Gallistel suggest that a fundamental problem with the Piagetian perspective is that the 'definition of how the younger child differs from the older is given in terms of what capacity *the younger child lacks*' (p.3). They criticise this aspect of Piaget's work for two reasons. First of all, they do not believe that children can be said to lack a particular ability on the basis of only one test. Secondly, they believe it is much more interesting to answer the question on the basis that a child moves from having one kind of concept to another, rather than moving to a concept from nothing at all. Gelman and Gallistel suggest that without a picture of the kinds of concepts a child is moving *from*, we are restricted to a description of a child's progress as opposed to an understanding of the *process* of development (Gelman & Gallistel, 1978). Modern Piagetian frameworks, such as Glasersfeld's 'radical constructivism', take this criticism into account and take pains to describe the process by which children construct their concept of the world, including its mathematical content.

This investigation is aiming to describe differences in the ways in which children make transitions in conceptual understanding of mathematical concepts. There is a limit to what can be learned by using a methodology in which analysis is restricted to static states. Alternative methods for data collection are considered later in section 5, as greater detail regarding the processes involved in children's problem solving becomes necessary.

2.8.2 The 'Experimental Contract'

In choosing to use experimental methods for this type of investigation, some issues are raised. The crux of these issues rests on the need to understand how participants perceive their involvement in the research. There is an assumption that participants, when given a set of problems to solve, are motivated to solve those problems to the best of their ability. This is not necessarily the case. For example, participants might believe that they do not need to try as hard in an experiment, as their teacher will not be seeing their work. Alternatively, participants may feel under pressure due to the unfamiliar conditions of a controlled experiment and therefore want to finish as quickly as possible. When participants are solving problems on multiple occasions, additional problems may be created. Criticism of Piaget's classic conservation studies (Piaget, 1952), focussed on the fact that children often change their answer if asked the same question twice (Samuel & Bryant, 1984). Children interpret the experimental situation in this way because they are unfamiliar with this form of questioning. Normally, when children are asked the same question twice, with no other feedback, it is because the questioner is looking for a different answer the second time.

Efforts must be made by the experimenter to create a situation in which participants behave in a way in which they are expected to. Some measures were taken in the research presented here in order to avoid the kind of experimenter effects described above.

The progression of experiments is an important factor in avoiding the 'asking twice' problem. The liklihood that participants will behave in an unexpected or unpredictable way increases as the experimental situation deviates from their usual classroom experience. Initially, therefore, the experiment will be designed so as to replicate as closely as possible children's usual classroom experience. As a more detailed understanding is required, experimental conditions will need to change. However, results in later experiments can be compared with the those obtained in the intial, classroom-like experiment, as a check on validity.

A second precaution taken is to be open and clear with participants about the nature of the experimental procedure. In order to avoid misinterpretation on the part of the participant, the experimenter must be clear about what the purpose of the research is and what will be done with the answers given by the participant. It should be clear that participants' answers will not be used for any assessment, also that the researcher is interested not in how good the participant is at mathematics, but how they go about solving unfamiliar problems. The fact that a very similar set of problems will be set 1 week after the first set should be clear from the outset, as should the fact that no feedback will be given. These measures should make the experimental situation more transparent to participants. This in turn should cause participants' behaviour to be more transparent to the experimenter.

Despite these measures, it will be important to bear in mind the validity issues discussed here during analyses. Comparisons of results across schools and across classes will help to show that children are not behaving differently according to the curriculum or the teaching methods that they have experienced. Results will also be compared across studies to ensure that, as the experimental situation deviates from participants' usual classroom experience, behaviour does not change.

2.8.3 Sampling

Sampling is clearly required for the study as described above; it is not possible to conduct such a study with a whole population. Therefore, how should students be selected for participation in the studies conducted as part of this investigation? There are three ways in which this decision can be made. The first is on the basis of practicality. The second is on the basis of theory. The third is on the basis of typicality.

Taking practicality first, as this will reduce the number of potential participants dramatically – it is impractical to include students in schools that require a long journey from Nottingham. It is very practical to sample entire classes of students. Participants will initially all be asked to complete the same tasks in exam conditions, so for the sake of efficiency it makes sense to have complete classes of students taking part together. Depending on the results obtained during initial studies with complete classes, further work will involve clarification of any findings, possibly involving work with individual children or small groups. In order to gain access to children for the purposes of this research, it will still be desirable to work with complete classes of children, so as to minimize disruption to schools and departments.

What effect does the theory have on the choice of participants? In other words, for which potential participants are the implications of the study most important? There are two aspects of this study that might influence the selection of participants. The first is the problem content. The planned study will involve the same problems as used by Mevarech and Stern. According to the UK curriculum (DFES, 2001), children should have been introduced to the concept of a line graph to represent a changing quantity by the end of year 7, the concept of a compound measure by the end of year 9 and the concept of gradient also by the end of year 9 (See section 2.6 for further detail). The problems assume some knowledge of these, so it would seem appropriate to select participants at year 9 or above. The second aspect of the study that might influence participant selection is its potential implications. One implication of the study is with regard to a potential 'dual route' for mathematics at GCSE, which would mean children who achieve highly in maths would continue to study for GCSE mathematics, while children who do not achieve highly would study for a more practical qualification like a GCSE in numeracy. The conclusions of Mevarech and Stern suggest that this would make learning even more difficult for students that are struggling; they suggest that children are likely to develop a more transferable understanding of mathematics if the focus of their learning is on more abstract rather than more realistic problem sets. This suggests that students at the end of Year 9 would be suitable participants, as they are they students for whom it is most important for the study to be valid.

Each time the sampling frame is reduced in the ways like those described above, a threat to generalisability is introduced. Due to the limits described above, the population is students in year 9 in schools in Nottinghamshire. The method of sampling that will be used will be a form of cluster sampling, as only one or possibly two schools will be involved in any given study. Can anything be done to make the most of the remnants of generalisability that remain? One possible answer to this question is to consider typicality. It may well be that statistically, on account of a limited sample of participants, validity can only be formally demonstrated for one school – staff in other schools could legitimately argue that the results from such a study have no meaning for them. However, if the school in the study is very similar in key respects to other schools, the researcher will have at least some room in which to argue that the observed effects could be expected in these other schools. Therefore, it might be advantageous to select the most 'typical' school of those available, so as to increase the potential to argue generalisability to others. What is it then that might make a school more or less typical? Mathematics classes are more similar now that they might have been ten years ago as a result of the introduction of the National Numeracy Strategy. Children from what might be a large number of primary 'feeder' schools converge together in year 7 of a secondary school having had a similar experience of maths teaching and content. The number of feeder schools that a secondary school has might still be important however, when considering typicality. So might the number of students in the school, the size of the catchment area, the distribution of socio-economic class or ethnicity of students or the distribution of grades achieved at KS3 SATS or at GCSE level.

There are serious limits to the generalisability of much educational research as a result of poor sampling (Gorard, 2001), however, trying to select a more typical school in which to find participants for a study may mean that some argument can be made. Also, despite the restrictions described above that imply the use of only one or two schools for any one study, it will be possible to work with different schools for different studies and compare results in order to further assess generalisability across classes and schools.

2.8.4 Ethical considerations

The first issue discussed here is that of access to participants. This depended on the cooperation of a number of key people within each school involved. Initial contact was always with the head of the mathematics department, giving an explanation of what the research at a given stage in the research involved, why it was being done and what disruption there would be to lessons. In addition to the head of the department, agreement was sought from the head teacher and from the teachers of the classes that were potentially participating. Support from the classroom teachers was considered to be particularly important in order to establish cooperation with students. Staff at the school in which each study was conducted were informed as to the kind of information that would be needed for each student and what would be done with that information – also that information from each student will be secure, anonymous and confidential.

Issues involved in dealing with staff at a participating school are relatively simple. The question of how the study should be conducted with regard to the students is a little more complicated. The code of ethics of the BPS includes the following: "…the investigator must attempt to ensure that participants (including children) know of their right to withdraw. When testing children, avoidance of the testing situation may be taken as evidence of failure to consent to the procedure and should be acknowledged." (British Psychological Society, 2000). There is a question of whether, when children are being asked to do what they might normally be asked by their classroom teacher to do in a maths lesson, the right to withdraw is appropriate. In a normal classroom situation, students do not have the right to withdraw from the work they are set. Classroom teachers are *in loco parentis,* and therefore can make the decision as to whether children can take part in research in the classroom. It is expected that in practice, decisions as to whether to inform students of the right to withdrawal will depend in part on negotiation with the head of the department and with relevant classroom teachers. However, it is desirable that children are able to feel that they have some ownership of the research and feel comfortable with their participation. With that in mind, then the course of action for the research reported in this thesis was to give each participant a full explanation of the nature and purpose of the investigation in which they are taking part and to request each participant's consent before proceeding. Although not strictly necessary, as consent will already have been gained from the head of department and classroom teacher, this course of action was considered to make participants more comfortable with their involvement.

Students were informed of the reasons why the research is being conducted, what information will be required from them and what will be done with the information that is provided. This information will be provided in an accessible way, at a level commensurate with the participants' understanding. All data collected will be stored securely, in observance of the Data Protection Act, and will be accessible only by the lead researcher.

In all other aspects, the research reported here complies with the ethical standards of both the BPS (British Psychological Society, 2000) and BERA (British Educational Research Association, 2004).

3 Study 1

3.1 Introduction

Despite being exposed to the same lesson materials and problem stimuli, children often seem to develop their understanding of mathematical concepts in very different ways to one another. The aim of the investigation reported in this thesis was to explore reasons for differences in children's mathematical development.

The literature discussed in the previous sections suggests that strategy variability is a key factor in children's development of mathematical concepts. The evidence presented in section 2.2 suggests that periods of high strategy variability are associated with the incidence of cognitive change. In section 2.3 evidence was discussed that suggests that there are individual differences in strategy variability. It seems reasonable, considering these two sets of evidence together, to assume that differences in strategy variability might account for some of the variation in children's development of mathematical concepts. The position taken in this investigation regarding individual differences (see section 2.3) is that individual differences can be thought of as a summation of group differences. If some of the variance in strategy variability can be accounted for by one or more group factors, then some headway will be made in developing our understanding of difference in children's development of mathematical concepts.

One aim of this first study was to examine the nature of any differences in children's levels of strategy variability. Levels of strategy variability will be determined by analysing children's explanation for their answers to two sets of problems administered one week apart. It will be possible to determine how similar each child's two sets of strategies are to one another. Levels of strategy variability will be tested with regard to their association with gender, identified in section 2.4.3 as the most likely of those candidate variables discussed to show a relationship with strategy variability.

The problem set used was taken from Mevarech and Stern (1997) as described in section 2.6. These problems are chosen due to their combination of accessibility and novelty for the children involved in the study. In terms of accessibility, children have all the knowledge and understanding required to determine the nature of the problems and the kind of answer that is required. There is no notation or other mathematical construct involved in the problems that is unfamiliar to children at this level (see section 2.6 for an analysis of the place of these problems in the UK national curriculum). In terms of novelty, the children involved in this study will not have encountered problems like these before. The novelty in these problems stems from the fact that children will be asked to make judgements of rates of change. The children involved in this investigation will have had experience working with graphs like those presented in this study, but will usually have been asked to engage with them only to the extent of finding a y-value for a given x-value, for example.

In addition to their accessibility and novelty, these problem sets are suitable due to the number of strategies available to children in determining a solution. Mevarech and Stern (1997) described eight different strategies that children used and categorised them as ‘logical-mathematical’, ‘specific point explanations’, ‘non-mathematical’ and ‘other’. The breadth of strategies available to children for solving these problems will aid analysis of the variability of strategy use.

Mevarech and Stern (1997) used these problem sets in order to investigate the effect on context on the transfer of mathematical knowledge between problems. The problem sets come in two isomorphic versions. One version is labelled ‘sparse’ and presents the problem in an abstract way, with axes labelled ‘x’ and ‘y’ and lines labelled ‘line a’ and ‘line b’. A second version consists of the exact same set of problems, but places them in a realistic context; this version is labelled ‘realistic’. An example of a realistic context for these problems is a graph showing the income of two companies as both incomes increase over a number of years. Children were shown one version of the problems one week, followed by the other version one week later.

Mevarech and Stern (1997) showed that the nature of children’s transfer of knowledge across problem sets was different depending on whether children were shown the sparse set first, followed by the realistic set, or the realistic set first, followed by the sparse set. Their findings indicated that effective transfer, and therefore a higher rate of improvement, was more likely when children were presented with the sparse version first and concluded that the reason for this was that the sparse version forced children into using logical-mathematical explanations, which were easily transferable to the realistic context. Where children were first presented with the realistic problem set, they were able to use pragmatic reasoning that could not be used again effectively for the sparse problem set.

What was lacking from Mevarech and Stern (1997), at least as far as the goals of the current investigation are concerned, was an analysis of the individual differences in children's patterns of strategy use. It was clear from the results presented that not all children followed the pattern of strategy use described above. Therefore the present study allows for a combination of goals. These are to replicate the findings of Mevarech and Stern (1997) regarding context and extend that study with a more detailed analysis of children's patterns of strategy use. An attempt will be made to account for any individual differences in patterns of strategy use that are observed, with group factors. It is predicted that there will be evidence for individual differences in the patterns of strategy use shown by participants administered problem sets one week apart and that some of these individual differences will be accounted for by children's gender.

Fennema et al. (1998) reports that boys are more likely to use 'invented' strategies than girls, while girls are more likely to use rote-learned, algorithmic strategies in order to solve arithmetic problems. Gallagher et al. (2000) showed that boys are more likely to use unconventional strategies to solve unconventional problems than are girls. These examples suggest that boys should be expected to show greater strategy variability than girls in this experiment. Girls are predicted to be more likely to use a similar set of strategies in the second sessions as in the first, therefore showing low variability.

3.1.1 Pilot work

Some pilot work was carried out prior to the current study in order to confirm some of the assertions made above. A series of four focus groups were conducted in which children were asked to solve the problems intended for use in the main study. The children involved in the focus groups were all in Year 9 (aged 13 or 14 years old) and covered an achievement range spanning from Mathematics SAT level 4 to level 7, according to teachers' assessments and Key Stage 2 tests.

Children around a level 5 on the mathematics SAT showed that they had access to the problems, in that they understood the questions asked and the domain in which the answer was to be found. These children also used the widest range of strategies in order to answer the given problems. The strategies used by this group are given in Table 3. As can be seen in the table, some of these strategies, if used correctly, will generate a correct answer. The remainder of the strategies described in the table might sometimes give a correct answer (for example, sometimes the higher line happens to be the line with the greatest rate of change), but will not provide a correct answer on the majority of occasions.

The main aim of these focus groups was to ensure that the sample of participants chosen for the study were at a point in their development at which they had access to the problems, but not a well formed understanding of how to solve them. A second aim was to classify the strategies used by children to solve the problems.

Table 3: Strategies observed during pilot work for Study 1

	Strategy	How this strategy is used
Suitable strategies	Relative gradient	Whichever line is steepest at a given point on the x-axis has the greatest rate of change at that point
	Calculation	Whichever line increases the most between two points on the x-axis has the greatest rate of change
Unsuitable strategies	Relative height	Whichever line is highest at a given point on the x-axis has the greatest rate of change at that point
	Points line up	The marked points on each line are aligned vertically, so the lines have the same rate of change
	Individual point	If the lines cross at some point, then up to that point, the lines must have been increasing at the same rate
	Relative length	Whichever line is the longer (or shorter) has the greater rate of change

Regarding the achievement level of participants suitable for inclusion in the sample for the study, it was found during the focus groups that a number of children below SAT level 5 had some difficulties in understanding the demands of the problems.

These children had a limited understanding and facility with graphs, unable in many cases, for example, when given a value on the x-axis, to respond with the corresponding values on the y-axis for a given graph. Where an answer was given, these children generally used the relative heights of lines in order to compare rates of change.

Children at a SAT level of 6 or above, on the other hand, generally showed a good understanding of the problems and very often responded with correct answers. These children generally used the relative gradients of the lines in order to make judgements about rates of change.

The main conclusion drawn for the pilot work was that children with an achievement level around an SAT level 5 would be suitable participants for this study as they showed the widest variety of strategies for the problems under investigation. As the main aim of the study is to investigate differences in strategy variability, this variety of available strategies is important.

3.2 Method

3.2.1 Design

There were two aspects to the design of this experiment. The first aspect was a replication of Mevarech and Stern (1997). The second aspect was an extension of that study, intended to explore patterns of strategy use across trials amongst participants, particularly focussing on differences in levels of strategy variability associated with gender.

With respect to the replication of Mevarech and Stern (1997), a two-way factorial design was used, the independent variables being problem type and problem order, with repeated measures on problem type. The dependent variable was the number of correct answers given, out of a possible total of 9, for a set of problems. There were two conditions for each independent variable. Problem types were 'sparse' and 'realistic'; Problem orders were 'sparse to realistic' or 'realistic to sparse'.

Predictions were that students would perform better on problems set in sparse context than in realistic context (problem type), and that greater improvement would be achieved by students moving from sparse to realistic context.

With respect to the investigation of strategy variability, a correlational design was used, in which the likelihood of strategies being used in the second trial, given their use in the first trial, was compared across genders. It was predicted that there would be a higher correlation between strategies used in the first and second sessions for girls than for boys.

3.2.2 Tasks

Three isomorphic sets of problems were used, adapted from the study of Mevarech and Stern (1997). Each task took the form of two printed A4 sheets stapled together. The top half of each sheet showed a graph – all of the questions in the set referred to the same graph, repeated on each sheet so that children would not have to keep turning back to the first page. In addition to the graph, on the first page there were some general instructions, which were also read to the children before they began writing answers to questions. The instructions advised children that they could do any workings out on the graphs if they thought it might help, and also that they should pay careful attention to their explanations when asked for.

On the second page there were 3 questions that were taken directly from Mevarech and Stern (1997), with only the wording changed (on the basis of the pilot focus group work described above) in order to improve children's understanding of the question being asked. These questions asked children about the rates of change of the two lines on the graph and also asked children to explain how they decided on their answer.

The only difference between the three sets of problems was the context. The sparse context problems involved a graph with axes labelled 'x' and 'y', and lines labelled 'line A' and 'line B'. There were two sets of realistic context problems; one involved a graph with axes labelled ‘income’ and ‘year’ and lines labelled ‘company A’ and ‘company B’, while the other involved a graph with axes labelled ‘amount of water’ and ‘time’ with lines labelled ‘tank A’ and ‘tank B’.

Examples of each of the problem sets used in this, and subsequent studies are given in Appendix A.

3.2.3 Participants

Participants were 45 year 9 students (aged between 13 and 14, 22 boys and 23 girls) studying at a school in Nottinghamshire. Students in the school were divided into two populations, or streams (X and Y). Within each stream, students were taught mathematics in sets from 1 to 5 according to achievement, where students in set 1 were the highest achieving. The participants in this study were all of the students present for both sessions in the two set 3 classes in the school (22 in class 9Y3, 23 in class 9X3).

The tasks were administered in the students’ usual mathematics classrooms. Students had studied line graphs in their classes, but had not formally studied the topics of gradient, compound measures or rates of change.

3.2.4 Procedure

Each complete class of students was assigned to one of two conditions. Class 9Y3 were asked to complete the set of problems with sparse context in the first session, then asked to complete a set of problems with realistic context in the second session. In the second session, one week later, students were divided into two groups; half were administered the changing level of water with time graph, half were administered the changing level of income with time graph. Class 9X3 completed the same tasks, in reverse order.

In order to determine group differences, mathematics SAT scores were recorded for the children taking part in the study. The Key Stage 3 SAT tests had been taken by the children 1 month prior to the beginning of this study.

3.3 Results

The analysis of results in this study consists of two parts. Firstly, there will be an analysis of the replication of Mevarech and Stern (1997), concerning the effects of context type and context order on test scores. Secondly, there will be an analysis of children's levels of strategy variability across sessions, focussing on associations with gender.

3.3.1 Effects of context type and context order on test scores

Before looking at students' performance on the problems, mathematics SAT scores were compared. The means for the two groups of students were 91.18 (Realistic to Sparse group) and 58.18 (Sparse to Realistic group, 1 absentee). A t-test showed a significant difference between the SAT scores of the two groups ($t=8.2$, $p<0.001$).

Correlations were calculated between mathematics SAT score, scores out of nine on sparse and realistic problem sets and amount of improvement between problem sets (See Table 4). Significant positive correlations were found between Sparse and Realistic problems ($r=.316$, $p<0.05$) and between SAT score and Sparse problems ($r=.353$, $p<0.05$).

Table 4: Correlations between SAT score, scores on sparse and realistic problem sets and improvement between problem sets

	SAT	Sparse Problem Set	Realistic Problem Set	Improvement between problem sets
SAT	1	0.353 *	-0.23	-0.127
Sparse Problem Set	0.353 *	1	0.316 *	-0.185
Realistic Problem Set	-0.23	0.316 *	1	-0.122
Improvement between problem sets	-0.127	-0.185	-0.122	1

* = Correlation is significant at $p<0.05$ (2-tailed)

Means and standard deviations of test scores for the sparse and realistic problem sets are given in Table 5.

Table 5: Means and Standard Deviations of scores (out of nine) for Sparse and Realistic problem sets (n=45)

	Mean	Standard Deviation
Sparse Problem Set	6.44	1.878
Realistic Problem Set	6.73	1.468

There was no significant difference between the means for the two problem sets ($df=44$, $t=0.977$, $p>0.05$).

Table 6 shows means and standard deviations for the degree of improvement shown by participants, according to context order. There was no significant difference in level of improvement between sessions due to context order (df =43, t=1.012 , p>0.05).

Table 6: Means and Standard Deviations of improvement between sessions, by context order

	Mean	Standard Deviation
Sparse → Realistic (n=23)	1.17	1.775
Realistic → Sparse (n=22)	0.64	1.787

During the analyses described in this section, it became apparent that boys showed a greater level of improvement between trials than did girls $F(1,44)=4.099$, $p=0.042$. This led to the supposition that there could be something different about the ways in which boys and girls approached these problems that caused the difference in level of improvement. The analysis of the relationship between gender and strategy variability was therefore a natural next step.

3.3.2 Strategy variability and gender

The quality of data obtained during this study regarding the strategies used to solve problems was not as good as had been hoped for. It was possible to determine whether children had used strategies involving the relative gradients of lines or not, but no finer grained analysis was possible.

Correlations were calculated between the incidence of the use of relative gradient in the first trial and second trial (see Table 7). Children were considered to have used a strategy involving relative gradient if either of the words "steep" or "steepness" appeared in their explanations for their answers. No child used the word "gradient" in their explanations.

Table 7: Correlations between the incidence of the use of steepness in the first trial and second trial

	Pearson Correlation Coefficient	Significance (2-tailed)
Girls	0.901	.000
Boys	0.439	.036

The correlations showed a marked difference between boys and girls, in terms of the likelihood of using strategies involving steepness in the second trial given their use in the first trial. Fisher transformations showed that the difference between the correlations observed for boys and for girls was significant ($z=3.14$, $p<0.05$). For girls, a test of linear regression shows that the incidence of an explanation involving relative gradient in the first trial predicts the incidence of an explanation involving relative gradient in the second trial ($F(1,44)=40.05$, $p<0.0005$). Conversely, for boys, an explanation involving relative gradient in the first trial does not predict an explanation involving relative gradient in the second trial ($F(1,44)=2.123$, $p=0.160$). 12 of 22 girls gave explanations involving steepness in the first trial. Roughly the same proportion of boys, 11 of 23, gave explanations involving steepness in the first trial.

Unfortunately, as discussed above, the data regarding strategies used were not of sufficient quality to test for correlations between strategies used in each session other than relative gradient.

Neither boys' nor girls' explanations were affected by the order of the problem contexts. As many explanations involving relative gradient were observed in first trials for sparse-context problems as for realistic-context problems.

3.3.3 Revisiting context order

As a gender difference in approach to these problem sets had been suggested by the analyses above, the data were analysed for the original hypothesis again, that improvement would be predicted by order of problem contexts; this time analysing data from boys and girls separately.

For the girls, the order of problem contexts does not predict improvement between trials $F(1,44)=0.269$, $p=0.609$. For the boys, there is some evidence to suggest that order of problem contexts does predict improvement between trials $F(1,44)=3.064$, $p=0.095$.

3.4 Discussion

There are three main areas for discussion on the basis of the results given above. Firstly, the lack of support for the findings reported in Mevarech and Stern (1997) requires some attention. Secondly, the findings concerning a potential association between gender and strategy variability will be discussed.

Finally, the use of the results of this study in informing the remainder of this investigation will be considered.

3.4.1 Comparisons with Mevarech and Stern (1997)

There was no evidence here to support the claims of Mevarech and Stern (1997). Context order had no effect on levels of improvement of the sample and there were no differences in scores on the sparse and realistic problem sets.

There were issues with the data generated in this study that may have contributed to the failure to replicate. The levels of achievement shown by the two groups of children, measured using Year 9 SAT scores, were such that the realistic-to-sparse group outperformed the sparse-to-realistic group by some margin. The scale of this difference was unfortunate. The two groups were drawn from the same school and were whole classes judged by their teachers to be at SAT level 5. The two classes were both Set 3 of 5 (set according to ability/achievement in mathematics with set 1 being the most able group of children) in parallel streams. The SAT scores were only available after other data collection had been completed, so the scale of the differences in level of achievement was not apparent until all data collection involving the problem sets had been completed.

The discrepancy in achievement level between the two groups may well have contributed to the failure to replicate the effect shown in Mevarech and Stern (1997). The correlations given in Table 4 suggest that differences in SAT score had some effect on participants' ability to solve the problems in at least one of the two problem sets.

3.4.2 Gender and strategy variability

There were fairly strong suggestions from the data reported here that gender is a good predictor of variability for the problem set under investigation. Girls seem more likely than boys to use a similar set of strategies for the second trial as used in the first trial. Boys are more likely to use a new set of strategies for the second trial. Of course, in this study it was not possible to conduct a very fine-grained analysis of the strategies used by children to solve these problems. However, there is sufficient evidence here to warrant a more detailed investigation of any association between gender and variability in subsequent studies.

In this respect, this first study should be considered a success. One aim of the study was to assess the role of gender in helping to explain individual differences in patterns of strategy use. As gender has been identified in this section as having a probable association with strategy variability, the following section will contain a detailed examination of the research literature relating to the effect of gender on strategy use in mathematics.

3.4.3 Informing future study design

More work is clearly required at this point in order to clarify the findings presented here. A second study with the aim of confirming and clarifying the findings of the initial study will be described shortly. The lack of detail in terms of the strategies used by children to solve problems was the major impetus for the design of the second study. The explanations given by children for their answers were enough only for a fairly surface-level analysis of strategy use. Some children gave good explanations for their answers, but a number of them gave either too little information to satisfactorily categorise their use of strategy, or no explanation at all. More detail regarding use of strategy will help to answer the main question arising in this section, concerning a more precise picture of the relationship between gender and strategy variability. In order to collect more detailed information on the strategies used by the children, they will be asked to solve problems individually. This will go some considerable way towards solving those problems described above that were encountered in the present study.

A second issue that will be considered for the design of the second study will be the matching of participants according to ability. The fact that one group of children were more able according to SAT score is not a problem for the analysis of differences between boys and girls as there was an equal proportion of boys to girls in each group. However, the difference in ability between groups did mean that analysis of context-order effects was unreliable. Due to the fact that SAT score and problem-set score were positively correlated, it is not clear whether there would have been a significant effect of context order had the two groups shown equal SAT scores.

3.4.4 Summary

There were some significant difficulties in analysing the data generated during this study. However, the problems encountered were relatively simple issues to rectify during future work, and some useful findings were obtained.

There was a suggestion of a link between gender and strategy variability, whereby boys show a higher level of variability than do girls. A replication of the results was required, as was further clarification of the effect through a more detailed analysis of boys' and girls' uses of strategy across sessions.

4 Study 2

4.1 Introduction

4.1.1 Strategy Variability

There has been an increasing amount of research over the last decade that has focussed on strategy change – the way in which children make the transition from a state where inaccurate and inefficient strategies are used to solve problems within a particular domain to a state in which accurate and efficient strategies are used. This transition can be described using the overlapping waves theory (Siegler, 1996), which represents the fact that children make use of a variety of strategies over a prolonged period of time, with the frequency of use of each strategy changing over time to favour the use of more accurate and more efficient strategies. In support of the overlapping waves model, it has been shown that children show a high level of variability within the individual, to the extent of showing variability of strategy use for the same problem presented on consecutive days (Siegler & Shrager, 1984). Children can even show variability of strategy use within a single trial. Alibali and Goldin-Meadow (1993) showed that when children are solving mathematical equivalence problems (e.g. 5 + 3 + 4 = ? + 4), they often show different uses of strategy in their verbal explanations of solutions as they do through their gestures.

Variability of strategy use has been shown to be an important predictor of performance, with high levels of variability at the point of introduction to a task being associated with high levels of later performance (Church & Goldin-Meadow, 1986; van der Maas & Molenaar, 1992; Siegler, 1995).

While there has been a large amount of research regarding typical patterns of strategy change, there has been little that has investigated individual differences. Differences in behaviour during periods of strategy change are interesting as they can help to answer questions regarding differing levels of performance in mathematics. One way of opening up questions of individual differences is to identify group factors associated with differences in behaviour. As will be shown below, gender is a group factor that is likely to be related to differences in strategy development.

4.1.2 Gender Differences in Mathematics Education

The literature on gender differences in mathematical cognition has been steadily moving from the general to the more specific over the past two decades. Hyde, Fennema, & Lamon (1990) found that a small difference in terms of achievement favouring males emerged in high school and college. They also found that the difference was greater among higher achieving students. Hedges and Nowell (1995) found a similar pattern of achievement, but also investigated the ratio of boys to girls among the top 10% of scores in standard mathematics tests. They found that among these top-scoring students, there were more boys than girls. Interestingly, Hedges and Nowell also note that while the gender difference in achievement in mathematics has been decreasing over the last several decades, the difference in the numbers of boys and girls amongst the highest achievers has remained constant. These performance findings are relevant for this study in the sense that they demonstrate that there is evidence for robust yet unexplained differences between boys' and girls' mathematics learning.

Research investigating the reason for any gender differences in achievement has been more recent. Researchers have begun to analyse gender differences in the processes involved in mathematics learning in order to answer questions regarding the differences in outcomes described above. Some studies have linked differences in achievement to differences in attitudes towards, or anxiety related to, mathematics (e.g. Ashcraft & Kirk, 2001; Nosek, Banaji, & Greenwald, 2002; Skaalvik & Skaalvik, 2004; Vermeer, Boekaerts, & Seegers, 2000). Others have associated differences in achievement with differences in ability in 'math-fact retrieval' (Royer, Tronsky, Chan, Jackson, & Marchant, 1999). There is also a growing literature concerned with gender differences in strategy use. For example, Fennema, Carpenter, Jacobs, Franke, and Levi (1998) asked children in their first years at school to solve simple addition and subtraction problems and showed that although there was no difference between boys and girls in proficiency and ability, there were differences in the kinds of strategies used to solve simple addition and subtraction problems. Girls tended to use traditional/taught strategies while boys tended to use more invented strategies. Carr has shown that boys in their first year of school tend to use retrieval rather than algorithmic strategies when solving arithmetic problems due to their emphasis on the social impact of strategy choice (Carr & Jessup, 1997). Gallagher and de Lisi (1994) classified problems from the SAT-Math paper (a college entrance examination for students in the US) as requiring either a conventional or unconventional strategy for solution, unconventional problems requiring some degree of insight or intuition or the use of a familiar algorithm in an unfamiliar context. They found that boys were more likely to match strategy use with problem demands than were girls. In a later study, the results of Gallagher and de Lisi (1994) were replicated (Gallagher et al., 2000) and the authors concluded that, 'strategy flexibility is a source of gender differences in mathematical ability'.

The research described in this section shows that strategy use varies according to gender in response to specific problems. The present study goes further to investigate gender differences in patterns of strategy use by comparing strategy decisions across problem solving instances.

4.1.3 Development of the Present Study

Any individual differences found in strategy development or variability would be important for a number of reasons. Siegler (1987) has shown that it is important not to rely on a single snapshot of data when investigating children's strategy decisions. The way that a child solves a problem in one instance does not necessarily tell the researcher very much about how that problem has been attempted in the past or how it might be attempted in the future. Averaging data across either children or trials can therefore be misleading. In the same way, if qualitative differences are found amongst the ways that children develop strategies, then it will be important for future work in strategy development to take account of these differences and to avoid misleading discussion of the "average" or "typical" child. Findings in this area could also have some impact on the classroom as they are likely to influence teachers' understanding of the need for differentiation in mathematics teaching.

Given the research discussed above, it seems that gender differences should be expected in the ways that children apply strategies across problem situations – the way that children transfer knowledge across contexts. An experimental method used in Mevarech and Stern (1997) provides a means of investigating such differences. One of the experiments reported in Mevarech and Stern involves children working on a set of problems related to rates of change of lines on a graph. There are two, isomorphic, sets of problems. One presents graphs that represent realistic situations; another presents graphs in a more abstract form. Half of the children in the experiment were presented with the realistic problem set followed by the abstract set one week later. The other half of the children received the sets in the reverse order.

The conclusions of Mevarech and Stern were that the order in which contexts are presented to students had an effect on students' improvement in understanding between trials. Students showed greater improvement when moving from an abstract version of the problem set to a more realistic version. Their interpretation of this finding was that the abstract version of the problem forced children to find a solution using a logical-mathematical strategy, easily transferred to the realistic version, while the realistic version could be attempted using a more practical, context-based strategy, not easily transferred to the abstract problem set.

Mevarech and Stern were using this method to investigate differences in transfer effects depending on context. However, the method seems equally appropriate for the purposes of the present study. The problems used lend themselves well to an investigation of patterns of strategy use. There are many possible strategies available to children, leading to both correct and incorrect answers. Also there is a clear distinction between reading off points from a graph and making judgements about rates of change, meaning that a wide range of children will have access to these problems without already having been taught strategies for finding solutions.

There was no analysis of differences within each group in Mevarech and Stern. For the present study it was predicted that for boys, the conclusions of Mevarech and Stern would hold mostly true. The evidence for boys' greater variability of strategy suggests that they might be more likely to be influenced by the context of a problem. For girls, it was predicted that students would be less affected by the context of the problem and would be more likely to show a lower level of strategy variability and therefore apply similar strategies to each problem set regardless of order.

In summary, the present study aims to use the experimental method of Mevarech and Stern (1997) to investigate differences between boys and girls in terms of patterns of strategy use across two sessions.

4.2 Method

4.2.1 Design

There were two aspects to this study. The first was the replication of Mevarech and Stern (1997), as in the previous study. To this end, a two-way factorial design was used, the independent variables being problem type and problem order, with repeated measures on problem type. The dependent variable was the number of correct answers given, out of a possible total of 9, for a set of problems. There were two conditions for each independent variable. Problem types were ‘sparse’ and ‘realistic’; Problem orders were ‘sparse to realistic’ or ‘realistic to sparse’. It was predicted that improvement between trials would be greater for those participants who saw the sparse version of the problem set prior to the realistic problem set that for those who saw the problems in the reverse order.

The second aspect to this study was the further investigation of differences in strategy variability according to gender. For this, a correlational design was employed, whereby comparison could be made of strategy sets used by participants in each problem solving session. It was predicted that boys would show a higher level of strategy variability than would girls, demonstrated by lower correlations of strategy sets used across sessions.

4.2.2 Participants

Participants were 58 13-14 year old children (24 girls and 34 boys) from two complete classes, one each from two schools in Nottinghamshire. All of the children were expected to achieve level 5 in the Year 9 SAT paper, the standard mathematics test taken by 14 year-olds in schools in the UK. Possible grades in the Year 9 SAT range from level 3 to level 8, and level 5 represents an approximately average level of performance. Pilot studies had shown that children at this level of attainment had access to the problem sets (children understood the questions and could come up with reasonable explanations for their answers) but had not yet developed stable strategies for finding answers. The children had not encountered rate of change problems in the classroom before.

The children were initially asked to complete Raven's Standard Progressive Matrices (Raven, 1976). This test is widely used to assess the non-verbal intelligence of children aged 8-16. There are 60 items in total, divided into five groups of 12. Each consists of a pattern in which there is a missing part. Children are required to select the correct part from a number of alternatives.

Lists for both boys and girls were drawn up in order of Raven's score. Alternate children were then placed in group 1 and group 2. The result of this process was four groups of children labelled 'male group 1', 'male group 2', 'female group 1', and 'female group 2'. The four groups did not differ significantly in terms of Raven's scores.

4.2.3 Tasks

Three isomorphic sets of problems were used, adapted from the study of Mevarech and Stern. These were the same as the problems used in the previous study. All problems involved children making judgements about rates of change on the basis of linear graphs. The instructions advised children that they could do any workings out on the graphs if they thought it might help, and also that they should pay careful attention to their explanations when asked for.

There were six questions, three each associated with two different graphs, taken directly from Mevarech and Stern (1997) with only the wording changed in order to improve children's understanding of the questions (based on previous pilot work). Each graph showed two straight lines that crossed at some point. These asked children about the rates of change of the two lines on the graph and also asked children to explain how they decided on their answer. For example, one question asked children, "after 1984, did the income of company A increase faster, slower or at the same rate as the income of company B?" Another question asked, "in 1984, was there a change in the rate of increase of income of company A?"

The only difference between the three sets of problems was the context. The sparse context problems involved a graph with axes labelled 'x' and 'y', and lines labelled 'line A' and 'line B'. There were two sets of realistic context problems; one involved a graph with axes labelled 'income' and 'year' and lines labelled 'company A' and 'company B', while the other involved a graph with axes labelled 'amount of water' and 'time' with lines labelled 'tank A' and 'tank B'.

4.2.4 Procedure

Children were administered the two tests, one week apart. Children in group 1 were initially given the sparse task, with the realistic task a week later. Children in group 2 were given the tasks in the reverse order.

Tasks were administered individually, taking approximately 15 minutes to complete in total. Each participant was asked to complete a set of problems. Once the set of problems was completed, the experimenter checked the explanations given by the children to ensure that enough detail had been given in order to determine the strategy employed.

The experimenter had the opportunity to ask each child for more information at the end of each session in order to elicit missing answers and to obtain further information regarding strategy. Questions were asked with the intention that children should not be led towards one explanation or another and were as neutral as possible, such as "could you tell me a bit more about how you did this one". The experimenter's responses were also as neutral as possible, intended to give no indication as to the correctness or otherwise of answers or of explanations.
Given that the author was solely responsible for coding strategies used by participants, it could be argued that there may be issues concerning the reliability of this coding. However, the experiment was designed specifically in order to address this issue. Where there was any doubt as to the strategy used, the experimenter had the opportunity to question the participant further in order to determine the strategy used. It is acknowledged that, even so, the recruitment of a second coder in order to generate a measure if inter-rater reliability could still be considered desirable. This was just not considered practical in this case. A similar posiiton is taken for this issue for each of the studies in this thesis.

4.3 Results

4.3.1 Distribution of strategies used

The strategies used by children in this study, with their distribution across instances of strategy use, are shown in Figure 2. The set of strategies used was similar to that observed previously, in study 1 (section 3) and in pilot work (see section 3.1.1). For a description of each strategy, refer to section 3.1.1. In this study, children were most likely to use explanations involving the relative height or steepness of the lines in order to solve the problems.

There were some differences in distribution of strategy use between boys and girls.

Figure 3 shows the distribution of strategies used broken down according to gender. The graph shows that girls were more likely than boys to use the relative heights of lines and less likely to use their relative steepness when making judgments about rates of change (χ_2=15.36, df=5, $p<0.01$).

Figure 2: Graph to show distribution of instances of strategy use (study 2)

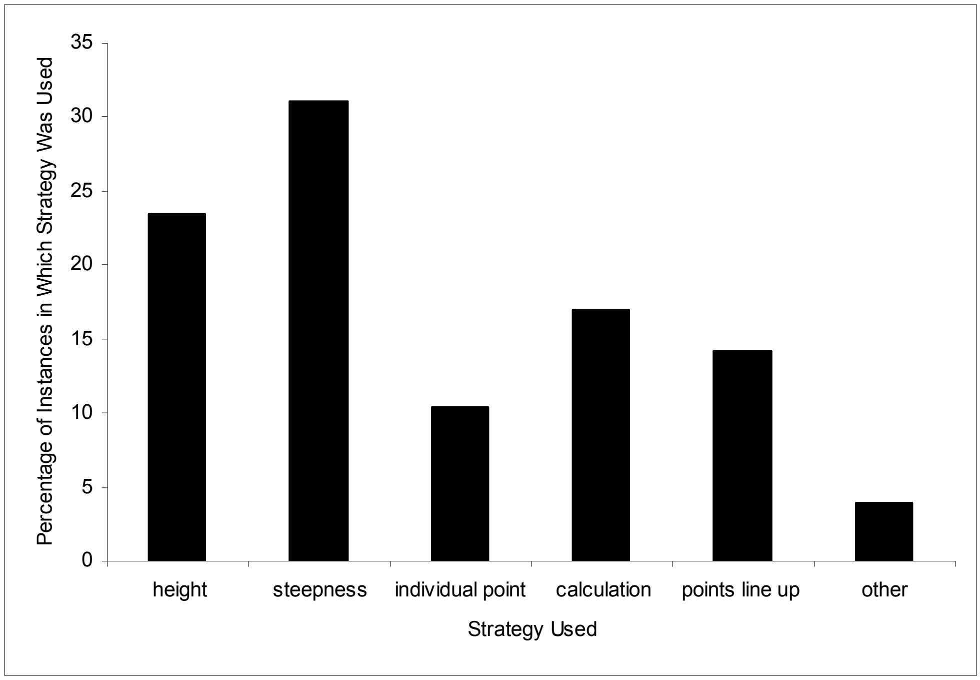
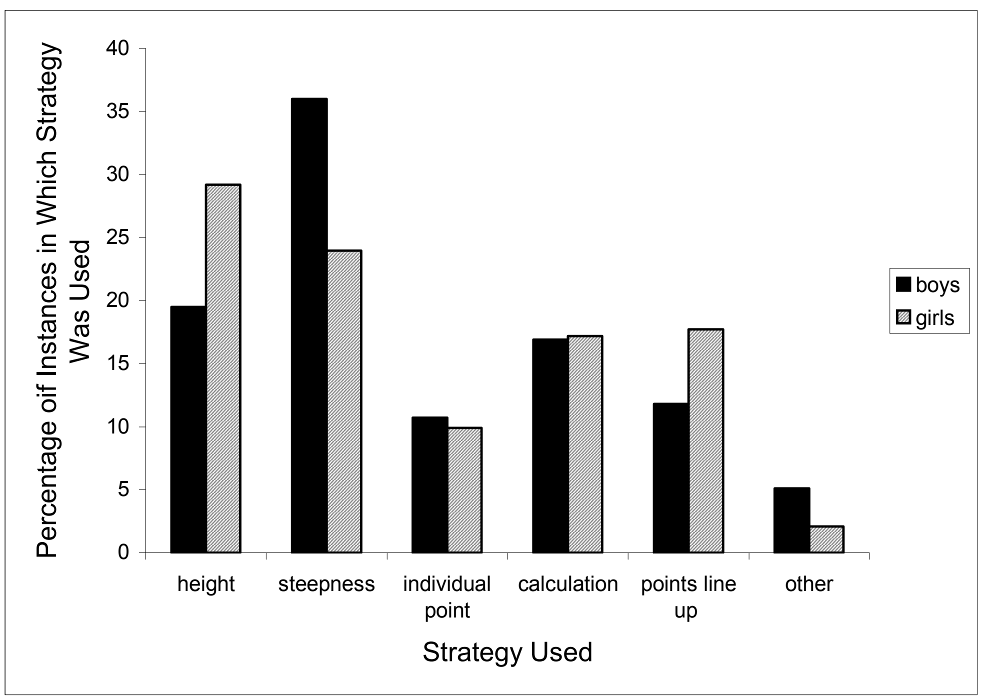


Figure 3: Distribution of strategies used, by gender (study 2)



4.3.2 Strategy variability

To assess the level of variability of strategy use of boys and girls, correlations were taken between children's use of strategy in the first and second sessions. This analysis was conducted for the two most commonly used strategies, the use of relative height and the use of relative steepness. For other strategies, there were too few instances of use in order to generate a reliable correlational analysis.

Table 8: Correlations between use of relative steepness in first and second sessions for boys and girls

	Pearson Correlation Coefficient	Significance (2-tailed)
Girls	0.504	0.012
Boys	0.180	0.308

Table 9: Correlations between use of relative heights in first and second sessions for boys and girls

	Pearson Correlation Coefficient	Significance (2-tailed)
Girls	0.694	<0.001
Boys	0.059	0.742

Table 8 and Table 9 show the correlation coefficients for use of relative steepness and relative height across sessions for boys and girls. The higher correlations shown by girls suggest lower strategy variability compared to boys. Fisher's z transformations were used in order to analyse the difference between correlation coefficients for boys and girls. These tests showed that the difference between girls and boys was significant for both relative steepness ($z=1.32$, $p<0.05$) and relative height ($z=2.82$, $p<0.05$).

4.3.3 Effects of achievement level, gender and order of contexts on test scores

Confirmation was made that the groups were matched appropriately and that there was no difference in baseline ability (measured using Raven's Matrices) between either groups 1 and 2 or between boys and girls. A two-way analysis of variance showed that the main effect of gender was not significant, $F(1,57)= 0.535$, $p>.05$), and neither was the main effect of group, $F(1.57)=0.869$, $p>0.05$.

Table 10 shows means and standard deviations of test score by gender, group and problem type. A 2x2x2 ANOVA with test score as the dependent variable and independent variables sex, context order and problem type shows significant differences between groups. There were significant main effects of sex, $F(1,57)=9.802$, $p<0.05$, and context order, $F(1,57)=2.831$, $p<0.1$. An interaction between sex and context order, $F(1,57)=4.902$, $p<0.05$, shows that the context order effect is accounted for mostly by the boys and seems to be due to a low average score for those who saw the sparse version of the task first (this can be seen in Table 10).

A 2x2 ANOVA with improvement in score between trials as the dependent variable and independent variables sex and context order shows no significant differences between groups. The main effect of gender was not significant, $F(1,57)=1.38$, $p>0.05$. Nor was the main effect of context order, $F(1,57)=0.497$, $p>0.05$.

Table 10: Means and Standard Deviations (in italics) of scores by gender and group (Study 2)

	Trial 1		Trial 2	
	Sparse to Realistic	Realistic to Sparse	Sparse to Realistic	Realistic to Sparse
Male	3.88 *1.73*	5.06 *0.97*	4.59 *1.33*	5.41 *1.00*
Female	3.92 *1.32*	3.91 *1.30*	4.07 *1.19*	3.82 *1.94*

4.3.4 Patterns of strategy use

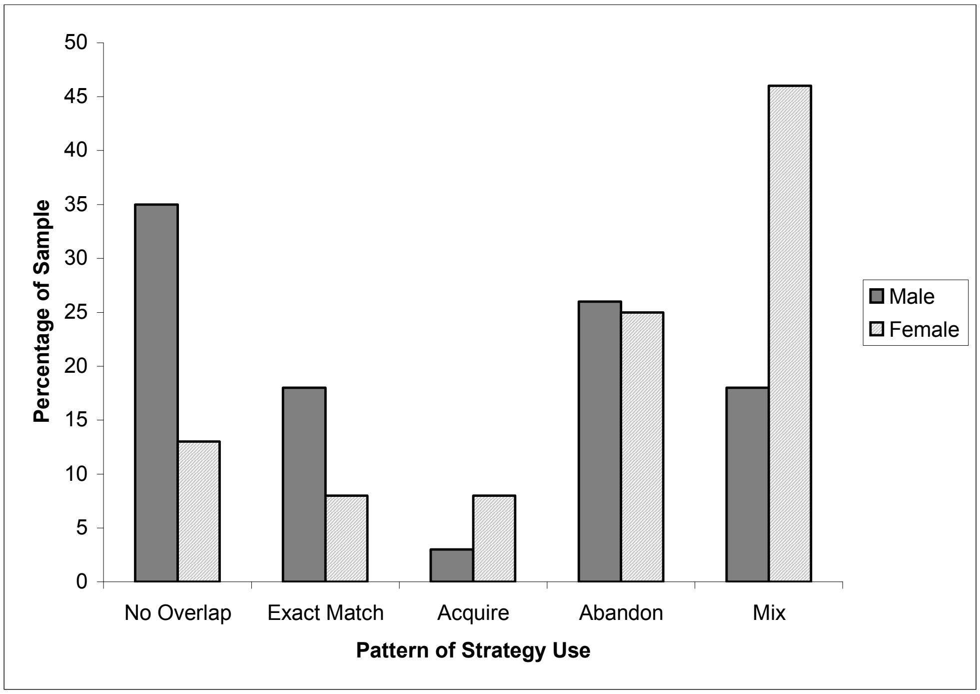
In order to investigate differences in patterns of strategy use between boys and girls further, each child in the study was labelled as showing one of five patterns of strategy use across sessions.

Compared with the set of strategies used in trial 1, in trial 2 children can use a completely new set of strategies (no overlap), use the exact same set of strategies (exact match), add one or more strategies to the set (acquire), lose one or more strategies from the set (abandon) or both add strategies to and remove strategies from the set (mix).

Figure 4 shows the distribution of boys and girls amongst these five categories. A χ_2 test shows that there is evidence that the distribution of participants amongst the five categories is different for boys than for girls (χ_2=6.78, p<0.1).

The main differences are apparent within the "no overlap" and the "mix" categories – where there are overrepresentations of boys and girls respectively.

Figure 4: Graph to show distribution of boys and girls across patterns of strategy use (study 2)



4.3.5 Summary

There were some minor differences between girls and boys and between context orders in terms of average score, and no differences in improvement between boys and girls or between groups across trials or differences in types of strategies used. However, there were sizable differences in terms of the patterns of strategy used demonstrated by boys and girls when comparisons were made across trials.

The key analyses in this study are those comparing children's uses of strategy across sessions. Correlational analysis demonstrates that for the two most commonly used strategies for this problem set, girls are more likely to use a strategy in the second session given its use in the first session. Closer analysis of children's patterns of strategy use showed that there were significant differences between boys and girls. For example, boys were approximately three times more likely than girls to use a completely different set of strategies in the second session compared to the first.

4.4 Discussion

4.4.1 Gender and Strategy Variability

Firstly, the association between gender and strategy variability will be discussed. The main aim of this study was to develop previous findings and establish the extent of any such relationship. The data generated here regarding strategies used by children were of a higher quality than in the previous study and allowed for a more reliable analysis of trends in use of strategy across sessions.

The results provide further evidence for the association suggested in section 3. There are large differences between boys and girls in the correlations between sessions of the use of the two most used strategies, relative height and relative gradient. Girls are more likely than boys to use one of these strategies in the second session, given its use in the first session.

These correlational analyses add to the growing evidence base for the existence of an association between gender and strategy variability. In section 2.3, a review of the literature suggests that high levels of strategy variability can indicate periods of conceptual change or times at which a child is ready to learn. It seems that a difference in strategy variability between boys and girls might indicate a difference in the way that boys and girls develop an understanding of mathematical concepts.

4.4.2 Patterns of Strategy Use

A second set of important findings to be discussed here are the differences in patterns of strategy use between trials according to gender. The data show a distinct difference between boys and girls in terms of the likelihood of using a particular strategy or set of strategies on the second trial given their use on the first trial. For both strategies for which there were sufficient instances for analysis (relative steepness and relative height) girls were more likely to use a strategy in the second trial given it's use in the first trial than were boys.

The literature cited in the introduction shows that there are differences in the ways that boys and girls make choices of strategies in response to mathematical problems. The findings presented here show that any gender difference in terms of strategy choice in response to individual problems is outweighed by a difference in terms of pattern of strategy use across problem situations.

Figure 4 shows the variety of patterns of strategy use demonstrated by children in this study. The distribution of participants amongst these categories fits expectations on the basis of the available literature. Alibali (1999), in a study of strategy change amongst 8 and 9 year old children solving equivalence problems, found that 24% of participants changed strategies abruptly (Alibali's "abrupt" change is equivalent with the present study's "no overlap" group), compared with 26% in the present study. Similarly, while Alibali found that 19% of her participants made no changes to their set of strategies between trials, in the present study 19% of participants fell in the "exact match" category. These similarities are striking considering the substantial differences in both population (8 and 9 year olds vs. 14 year olds) and domain (equivalence problems vs. rate of change problems).

The variety of patterns of strategy use is interesting because a key assumption of this study was that the children involved had not encountered problems of this kind before. Therefore during the first trial, they were generating a set of strategies (one or more) to solve the set of problems. None of the children received any feedback about the correctness or otherwise of either their strategies or their answers. Given then, that the experimental procedure was identical for all participants, excluding to a large extent effects of teacher action and social interaction, there is evidence to suggest that differences in pattern of strategy use are a function of some set of factors within each child. The difference in distribution across categories of patterns of strategy use between boys and girls suggests that gender is one of the factors within that set. The findings discussed here certainly add to our picture of the way in which students develop an understanding of mathematics. The analysis of strategy use across trials makes clear the fact that it can be misleading to homogenise groups of students in terms of their process of development or understanding.

Using the same experimental method as Mevarech and Stern (1997) it has been possible, with additional analysis of differences within experimental groups, to achieve an additional level of description of students' understanding. While further research would be needed in order to generalize this finding to children of other ages and abilities and to other topic areas in mathematics, these differences do fit into a pattern of existing research, adding some confidence to the findings.

The data can be interpreted within Siegler's (1996) overlapping waves model of strategy development. The literature in this area suggests that strategy development is cyclical in that children move between periods of high and low strategy variability, with conceptual change being associated with periods of high variability (e.g. Siegler, 1995; Alibali, 1999). If high variability is associated with conceptual engagement, then individual and group differences in strategy variability will be an important area for future research.

4.4.3 Gender and Context Effects on Test Scores

As in study 1, reported in section 3, the context effects reported by Mevarech and Stern (1997) were not replicated in the present study. There was not sufficient evidence to suggest an effect of context order on improvement. However, there was some evidence that boys were affected by context to a greater extent than were girls. For boys only, greater improvement was shown by the group who saw the sparse version of the task first than by the group who saw the realistic version first. This is similar to the finding reported in Mevarech and Stern (1997) for the sample as a whole.

In this study, the findings regarding differences in outcome do not seem as important as those regarding process. The observed differences in strategy variability and, more generally, in patterns of strategy use, add more to our understanding of children's mathematical development than any differences in test scores.

4.4.4 Informing later work

It is useful to know that there are two types of problem solving behaviour in a classroom. It may be, on the other hand, detrimental to say that there is a gender difference in problem solving behaviour without an understanding of why such a difference might exist. This is due partly to the fact that not all boys and girls fit their respective patterns, also that the differences in behaviour may at least in part be caused by teacher and peer expectations. It is almost certainly not appropriate to direct teaching in different ways to girls and boys within a class. Further work in this area is likely to include the investigation of those factors influencing children's strategy decisions in order to understand why it is that patterns of strategy use differ.

Both the participants and the problems used in this study represent fairly narrow samples of their respective populations. The similarities of the distributions of children amongst the various categories of patterns of strategy use in Alibali (1999) and in the present study goes some way to alleviate this concern, but clearly more work is need in order to consolidate these findings and generalise across age, ability and domain.

There are two ways in which the findings presented here can be extended. The first is to investigate additional individual and group factors that influence patterns of strategy use. The identification of gender as one factor associated with strategy development raises the question of what other group factors might also be associated in the same way. The second is to investigate further the nature and causes of the gender effect on patterns of strategy use. In the further exploration of the gender effect reported here it will be important to identify any covariants with gender that could explain differences in strategy development. The data in the present study seem to be compatible with some of the literature in mathematics education concerning the relationship between gender and affect. For example, some studies report differences in levels and effects of maths anxiety between boys and girls (Lussier, 1996; Miller and Bichsel, 2004). It would be interesting to see what proportion of the variance in strategy variability would be accounted for by children's level of math anxiety independent of gender. Another possible explanation of the observed gender effect might be found by investigating differences in risk-taking behaviour. There is a large amount of research in the literature regarding gender differences in risk taking (Byrnes, Miller, & Schafer, 1999). The data reported in the present study show that boys are much more likely than girls to completely abandon one set of strategies in favour of a new set for the second trial, while girls are more likely to reuse at least some of the strategies used in the first trial. The boys' behaviour seems to demonstrate a higher degree of risk-taking than does the girls'. The association between strategy development and risk-taking deserves further investigation.

4.5 Conclusions

The main finding of this study has been that there are differences in the ways that boys and girls choose strategies across problem situations. This indicates that it will be important that researchers do not ignore group and individual differences in strategy development. Caution must be taken when describing the strategy development of the "average" or "typical" child. The fact that there are gender differences in patterns of strategy use in the present study suggests that there is a great deal of work to be done investigating the extent to which differences due to gender, ethnicity, socio-economic status and other factors might influence strategy development.

5 The need for more detailed data collection

On the basis of the first two studies reported in this thesis, there is a need for more detailed data to describe the relationship between gender and strategy variability. It is clear that there is some association between gender and strategy variability, but the nature of that association is not clear. Gender effects, as discussed in section 4.4, are of limited use without further investigation. This is because there are often factors other than gender that can better account for cognitive differences than sex alone. For example, a great deal of work has been done over the last decade in an attempt to establish how much of the gender difference in mathematics achievement can be accounted for by factors such as mental rotation and anxiety. These findings of course provide a very much clearer picture of differences in children's processes of mathematical learning than the finding of a gender difference, and are more likely to be successfully put to use in the classroom. The problem with gender differences in cognition is that variation within a gender is almost always considerably greater than any variation between genders. Also, there are so many factors (physical, biological, social) that covary with gender, that it is impossible without further investigation to construct a model that can describe the cause or origin of those cognitive differences.

The next stage in this investigation involved some consideration of likely mediating factors, covarying with gender, that could help to explain the relationship between gender and strategy variability. In order to do this, it will be necessary to consider some additional data that can be collected during children's attempts to solve the problems under investigation. In this section, options for this will be discussed.

We have an idea then of the functions that the additional data will be required to perform. One of the more important goals in later studies will be to ensure that strategies used by children can be classified as reliably as possible. A second goal will be to explore the evidence for other factors' influence on strategy variability, in order to better understand the reasons why gender predicts strategy variability.

5.1 Verbal Protocols

In order to as accurately as possible determine the strategies used by children to solve problems, verbal reports are often used. These can take a few different forms. Verbal data can be collected during problem-solving, either by prompting children with questions or by asking children to think aloud the whole time that they are working on a problem (Ericsson & Simon, 1993; van Someren, Barnard, & Sandberg, 1994). Otherwise, data can be collected after children have completed a problem, via retrospective prompting.

There is a good amount of evidence to suggest that a combination of think-aloud protocols and retrospective reports is the best means for gathering data regarding children's uses of strategy when solving mathematical problems. The main advantage of collecting data while children are actually solving problems is to prevent the loss of information through children's memory failure when reporting problem solving procedures at the end of an activity (Wade, 1990).

The main disadvantage of collecting verbal data while children are working on problems in that it poses a risk of disrupting or altering the problem solving process that is being measured. Think-aloud protocols as described by Ericsson and Simon (1993) address both of these issues – they are claimed to provide data on children's cognitive processes while they are occurring without disrupting or altering those processes.

Kuusela and Paul (2000) report a study comparing the use of think-aloud protocols and retrospective reports in terms of effectiveness. They found that concurrent think-alouds outperformed retrospective reports in general, eliciting more statements and greater insight into problem solving processes. However, they concluded that the use of retrospective reports could be a valuable *addition* to think-alouds, as they provided more information about children's final decisions in the problem solving process and could provide useful support for the data elicited by the concurrent think-aloud protocols.

There are some limitations of think-aloud protocols that should be borne in mind during their use. One major criticism of verbal reports as data can be found in Nisbett and DeCamp Wilson (1977) in which it is claimed that participants do not have direct access to cognitive processes. Verbal reports of those processes will only be accurate where "influential stimuli are salient and are plausible causes of the responses they produce". Young (2005) lists three methodological issues involved in the use of think-alouds; reactivity, participants' verbal ability and validity in analysis. The problem of reactivity is that the use of think-alouds might have the effect of altering the problem solving processes that those protocols are intended to inform on. Stratman and Hamp-Lyons (1994) discuss a few causes of reactivity when using think-aloud protocols. For example, when participants in a study are asked to think aloud, they are not working on a problem in their usual way. They are being asked to complete an additional task, which may have an impact on their available cognitive resources. Participants are also having their attention drawn to the cognitive processes involved in the task set, to a degree which it would probably not in normal circumstances. Participants' verbal abilities can cause a problem due to the variation within a sample. Children can vary widely in their ability to provide a verbal report on their cognitive processes. Wilson (1994) and Branch (2000) both describe research situations in which the variation of the completeness of protocols has caused difficulties in interpreting the data. The main threat to validity in the use of think-aloud protocols is the fact that children are not necessarily able to verbalise the entire cognitive process that they experience when solving a problem. We know, for example, that there are unconscious aspects of strategy discovery and children can begin to use a strategy some time before they have the ability to verbalise an understanding of it (Siegler & Stern, 1998). Also, there are aspects of problem solving that are automatic and therefore unlikely to be reported (Wade, 1990).

It will be important to try to minimise the effect of those issues described above. This can be done in a few different ways. Firstly, the use of retrospective reports in combination with concurrent think-aloud protocols, as mentioned above, will provide a means to triangulate the data. A consistent story from both sources will increase validity, whilst conflicting information will indicate a need for caution in the interpretation of the data. Secondly, there must be an awareness of what is realistically hoped for with the use of the think-aloud protocols. It is not reasonable to expect a clear and complete description of a child's cognitive processes from a concurrent verbal report. The use of verbal protocols must be thought of as a means to enhance the data available for understanding children's uses of strategy in response to problems, and no more. In the studies to be reported later in this thesis, think-aloud protocols are used as a means to increase the reliability of strategy categorisation and to provide an indication of factors such as the nature of strategies other than the final strategy considered, and the level of certainty with which children decide on a particular strategy. They will not be used with the intent to generate a complete picture of children's cognitive processes in generating a solution to a problem.

5.2 Time on Task

The amount of time that children spend on task when solving a problem is often used in categorising strategy use. As a measure of children's problem solving processes, time on task has the advantage of being very objective. While it may not always be clear why exactly children are spending different amounts of time on a task than one another, it is certainly possible to say that if there are substantial differences between groups in the time it takes to solve a problem, then there are differences in their problem solving processes.

In the next study, time on task will be used in order to try and pick up any differences in participants' problem solving processes that might help explain the gender difference in strategy variability observed in the studies reported up to now. Analysis of time on task will come in two forms; differences between boys and girls can be tested for, as can differences between participants showing high and low levels of strategy variability. If there is some mediating factor that is associated with time on task that can help explain the relationship between gender and strategy variability, then a significant difference should be found in each test.

It will be important to bear in mind what such a finding could indicate. There are a number of factors that might lead children to spend more or less time on task. Some of the variables discussed in sections 2.5.3 and 2.5.4 are included, such as children's affective response to mathematical problems and their temperament. The strategy used to solve the problem may be an influence, although this would probably disguise any association with gender or strategy variability as these have been shown to be independent of the strategy used.

Time on task, then, will be used as something of a 'catch-all' test, intended to highlight any possible association that might require further investigation. The use of this data is limited in the sense that a significant result will not immediately provide an answers to the questions we have, but it is useful in the sense that it will certainly narrow the search.

6 Study 3

6.1 Introduction

The two studies described so far in this thesis have shown that gender is a good predictor of strategy variability for the problem set under investigation. Boys have shown higher levels of variability, compared with girls, and distribution of participants across categories of patterns of strategy use is different for boys than for girls. This finding is valuable and goes a long way towards achieving the major aim of this thesis, but also raises a number of questions.

The most important question is what is it about gender that is causing children to demonstrate differences in strategy variability? One possibility that needs to be investigated is that gender co-varies with some other factor that is the cause of differences in strategy variability. There are a few candidates for factors that might be playing this role. It will be necessary now to consider what factors, co-varying with gender, might be affecting children's decisions as to whether to use a similar set of strategies in the second trial as the first.

One strong possibility might be memory for either the problems encountered, or the strategy set used, during the first trial. Math anxiety is a second potential co-variant that might help explain differences in strategy variability. Other potential co-varying factors might include temperament and motivation.

It seems likely that children with a better memory for the problems and strategies would be more likely to use a similar set of strategies in the second trial to that used in the first trial, given that the strategy set gave rise to a favourable outcome in the first trial. There is some evidence that memory is a co-variant with gender (Herlitz, Nilsson, & Backman, 1997; Maccoby & Jacklin, 1974). Maccoby and Jacklin (1974) found that girls' memory is superior to boys', which could help to explain girls' lower level of strategy variability. Herlitz et al. (1997) also showed that girls performed at a higher level than did boys on tests of episodic memory. If girls were better able to remember both the problem set and the strategies they used to answer the problems in the first trial than boys, then they could be more likely to use the same strategies in the second trial.

The mechanism described above, however, could be complicated by the fact that math anxiety has been shown to have a negative affect of working memory performance, which in turn would affect long-term consolidation. Math anxiety has been shown to be a strong mediating factor of the gender effect in working memory performance, with girls being more prone to math anxiety than boys (Miller & Bichsel, 2004). Miller and Bichsel found that math anxiety affects visual working memory resources, which would clearly be in use when dealing with the problem set currently under investigation. It is not possible to assess working memory load during problem solving, as the problems under investigation are such that children's full attention is required. If children are asked to perform any other task concurrently with solving the problem sets, then problem solving will be disrupted to a great extent. It may be possible, though, to obtain some measurement of math anxiety from students' behaviour while working on the problem sets. Given what is known about the relation between math anxiety and working memory load, it would be possible to infer that children showing a higher level of anxiety during problem solving would likely have a reduced working memory capacity available for problem solving.

There is some evidence to suggest that differences in temperament can explain differences in strategy use between boys and girls. Davis and Carr (2001) showed that differences in temperament could help explain why boys and girls solved problems in different ways. Boys' strategy choices were associated with their level of impulsiveness, whereby boys with high levels of impulsiveness were more likely to use manipulatives to solve simple arithmetic problems while boys with low levels of impulsiveness were more likely to use retrieval strategies. Girls showed no differences in strategy use associated with impulsiveness, but instead showed differences according to their level of inhibition. Girls with higher levels of inhibition were less likely to use retrieval strategies. It will be possible to derive a measurement of temperament based on observations of problem solving behaviour including time on task.

Based on observations made during previous studies, an additional potential co-variant with gender will be considered. In the previous studies and pilot work reported in this thesis, it has been noted that children appear to vary in their perceptions of contradiction regarding potential strategy sets. A significant number of participants have used a strategy set in the second trial that contradicts the strategy set used in the first trial. The most common example of contradictory strategy sets used by participants is the use of strategies involving the relative height of lines and strategies using the relative steepness of lines. These two sets of strategies are contradictory, for many questions giving opposite answers. It seems possible that an understanding of contradiction would lead children to select compatible strategy sets. It will be possible, through questioning, to determine whether understanding of contradiction is associated with low strategy variability.

The aim of this study is to elaborate on the findings observed so far, to provide data that will identify potential co-variants with gender, particularly those described in this section, that might help explain the association between gender and strategy variability. This additional data will be accumulated through the use of think-aloud protocols and retrospective reports as discussed in previous sections.

6.2 Method

6.2.1 Design

In this study, the independent variable was gender and the dependent variable was strategy variability. The main prediction of this study was that there would be an effect of gender on variability, with boys showing greater variability than girls. Strategy variability is determined through comparison of children's uses of strategy in two problem solving sessions, one week apart.

In addition to this, a number of other factors are hypothesised to show an association with gender and with strategy variability. These are time on task, memory for the problem set from the first session, memory for the strategy set used in the first session and perceptions of contradiction amongst strategies.

Each of these factors hypothesised to mediate the gender effect on strategy will be treated as a dependent variable, with gender and strategy variability as independent variables. These tests will be conducted separately, as gender and strategy variability have already been shown to not be independent of one another.

6.2.2 Participants

Participants were 13-14 year old children from two complete classes, one each from two schools in Nottinghamshire. Both schools' Mathematics departments set children in terms of ability and both classes sampled were set 3 of 5, with set 1 being the most able. These students were selected in this way so as to match as closely as possible with participants in each of the last two studies, in order to limit sources of variation in behaviour.

The children were initially asked to complete Raven's Standard Progressive Matrices (Raven, 1976). Lists for both boys and girls were drawn up in order of Raven's score. Alternate children were then placed in group 1 and group 2. The result of this process was four groups of children labelled 'male group 1', 'male group 2', 'female group 1', and 'female group 2'. The four groups had approximately equal average Raven's scores.

6.2.3 Tasks

Three isomorphic sets of problems were used, adapted from the study of Mevarech and Stern. These were the same as in the previous study. Each task took the form of three printed A4 sheets stapled together. The top half of each sheet showed a graph – all of the questions in the set referred to the same graph. The instructions advised children that they could do any workings out on the graphs if they thought it might help, and also that they should pay careful attention to their explanations when asked for.

On the second and third pages were 6 questions, taken directly from Mevarech and Stern (1997), with only the wording changed in order to improve children's understanding of the questions, on the basis of pilot work. These asked children about the rates of change of the two lines on the graph and also asked children to explain how they decided on their answer. The 6 questions were in two groups of three, each set of three questions referring to one of two graphs.

The only difference between the two sets of problems was the context. The 'sparse' context problems involved a graph with axes labelled 'x' and 'y', and lines labelled 'line A' and 'line B'. There were two contexts used for the 'realistic' problems; one involved a graph with axes labelled 'income' and 'year' and lines labelled 'company A' and 'company B', while the other involved a graph with axes labelled 'amount of water' and 'time' with lines labelled 'tank A' and 'tank B'.

6.2.4 Measures

A number of measures were used during this study. These fall into three main categories. Firstly, base-line measures, intended to describe children's level of achievement or ability in school mathematics. Secondly, there were strategy measures, intended to represent children's strategy choices in response to problems. Thirdly, there were process measures, used to describe aspects of children's problem solving processes.

The base-line measures used for this study were Key Stage 2 Mathematics SAT score (taken at age 11), Key Stage 3 Mathematics SAT Score (taken at age 14) and children's score on Raven's Standard Progressive Matrices (Raven, 1976).

The Key Stage SATs are taken by all students in the UK at ages 7, 11 and 14 and are an indicator of performance across the National Curriculum in Mathematics. These scores were used in order to determine whether the association between gender and strategy variability was independent of performance or achievement in classroom mathematics.

The Standard Progressive Matrices is a test that is widely used to assess the non-verbal intelligence of children aged 8-16. There are 60 items in total, divided into five groups of 12. Each consists of a pattern in which there is a missing part. Children are required to select the correct part from a number of alternatives. This test was used as described in 6.2.2 in order to form matched pairs. Raven's SPM, as opposed to SAT score, was used for this for two reasons. Firstly, the Key Stage 3 SAT score was not available until after the data collection for the study had taken place, and the Key Stage 2 score was considered to have been obtained too long before the study to be a reliable predictor of current achievement. Secondly, Raven's SPM is more specifically targeted at non-verbal reasoning and pattern identification, thought to be a key component of graph problem-solving (Carpenter & Shah, 1998)

The strategy measures taken for this study were related to the strategies used by children to solve each of the six problems in each of the two trials. Strategies used to judge relative rates of change were expected, on the basis of previous studies (see section 3.1.1), to include:

- Relative gradient/steepness of lines
- Calculation of gradient
- Relative height of lines
- Fact that the points on the lines 'line up'
- Fact that lines meet at a point
- Relative length of lines

Strategies were classified using think-aloud data where possible. When it was not possible to determine a strategy on the basis of think-aloud data, the retrospective report was used. Where the participant considered multiple strategies, the first strategy leading to the given answer was recorded.

Process measures taken included data concerning the time taken to answer each of the six questions in each problem set, and also answers given to questions asked as part of the interview conducted at the end of the second trial. This interview will be discussed in the next section (6.2.5).

6.2.5 Procedure

Children were administered the two tests, one week apart. Children in group 1 were initially given the sparse task, with the realistic task a week later. Children in group 2 were given the tasks in the reverse order.

At the beginning of trial 1, each participant was set a minimum of 2 practice problems, used for think-aloud training. The example problems, and instructions for think-aloud training were taken from Ericsson and Simon (1993 p.376-7). Training lasted between 5 and 10 minutes per participant.

Tasks were administered individually, taking approximately 15 minutes to complete in total. Each participant was asked to read each question out loud and then think aloud while completing the set of problems. The experimenter interrupted only after a period of silence (15-20 seconds) to remind the participant to keep talking.

Once the set of problems was completed, the experimenter asked for a retrospective report for each of the six questions in the problem set. This questioning took the form of: “What was the first thing that you thought after reading this question? What did you think next?… and so on.

The experimenter had the opportunity to ask each child for more information at the end of each session in order to elicit missing answers and to obtain further information regarding strategy. Questions were asked with the intention that children should not be led towards one explanation or another and were as neutral as possible, such as “could you tell me a bit more about how you did this one”. The experimenter’s responses were also as neutral as possible, intended to give no indication as to the correctness or otherwise of answers or of explanations.

The procedure for trial 2 was similar, with the exception of the think-aloud training and the addition of a structured interview, after all questions and retrospective report were completed, designed to provide data regarding participants’ perceptions of the similarities and differences between the two problem sets. Questions used were:

- “Forgetting about your answers for now, what similarities and differences can you remember between the problems you have just worked on and the ones you saw last week?”
- Following this question – further prompts were used, such as “can you tell me a bit more about that?” or “is there anything else that you can remember?”
- If participants struggled with the question, prompts such as “is there anything different about the graphs…. about the questions?” were used in order to narrow the range of answers possible for the participant to give
- “Do you think that one or other of the two sets of problems was harder to answer?”

- “Are there any differences in the ways that you have answered the questions this week, compared to last week?”

Then if necessary, prompts such as “and have you used any methods this week that you didn’t use last week to answer the questions?” were used

At this point children were shown their completed problem sheets from trial 1 alongside their completed sheets from trial 2 and asked again about any questions where they had changed their answer from one week to the next. The isomorphic nature of the problem sets was explained, so that all participants understood at this point that although the questions were worded differently, and the graphs were labelled differently, they were essentially asking the same thing.

For each question where answers differed between trials, participants were asked:

- “Can you remember how you got this answer last week?”
- “Are you still happy with both of these answers? Do you think you would change either of them if you could?

Following all questions, participants were given the opportunity to ask any questions. Sessions lasted between 20 and 30 minutes in total.

6.3 Results

6.3.1 Distribution of strategies used

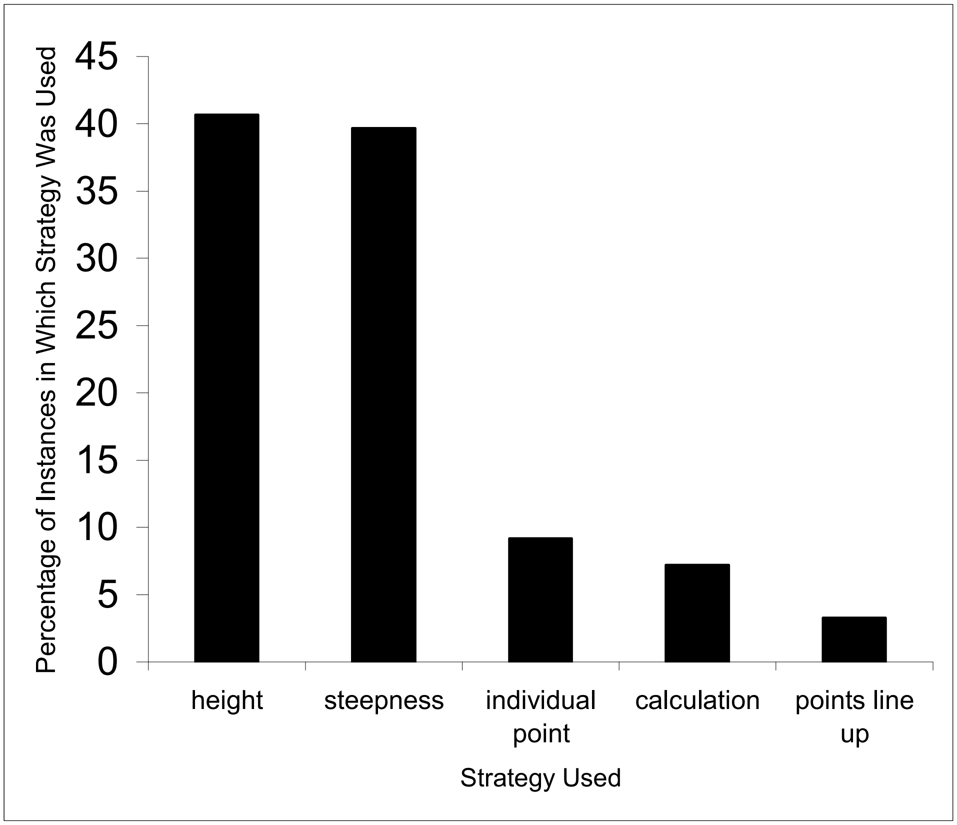
Table 11 shows the frequencies of strategies used according to session and gender. There is no difference between the distributions of strategies used in the first and second sessions (χ^2= 8.55, df=4, p>0.05).

Table 11: Frequencies of strategies used according to session and gender (study 3)

		Height	Steepness	Individual Point	Calculation	Points Line Up	Total
Session 1	Boys	17	29	6	2	2	56
	Girls	35	37	12	7	5	96
Session 2	Boys	22	24	3	6	1	56
	Girls	49	31	7	7	2	96

Nor is there a difference in the distributions of strategies used by boys and girls (χ^2= 4.38, df=4, p>0.05). Strategies used by children are also illustrated in Figure 5, with the proportion of questions for which each strategy was used. Strategies used were as predicted and the distribution was similar to that observed in study 2. However, in this study there was a more heavy weighting on ‘steepness’ and ‘height’ than in previous studies. Also, there was an absence of some of the anomalous strategies used in both previous studies, such as the use of the length of the lines on the graph.

Figure 5: Graph to show distribution of instances of strategy use (study 3)



6.3.2 Strategy variability

Repeating the correlational analysis of the previous study, correlations were calculated between use of strategies in the first and in the second trials for boys and for girls.

This analysis was carried out only for strategies involving the relative steepness of lines and the relative heights of lines, as there were too few examples of other strategies for any useful analysis (see Table 12 and Table 13).

Table 12: Correlations between use of steepness in first and second trials for boys and girls

	Pearson Correlation Coefficient	Significance (2-tailed)
Girls	0.176	0.410
Boys	0.623	0.017

Table 13: Correlations between use of height in first and second trials for boys and girls

	Pearson Correlation Coefficient	Significance (2-tailed)
Girls	0.120	0.578
Boys	0.713	0.004

Fisher transformations for both strategies analysed in this way showed boys' strategies to be more highly correlated than were girls' (for steepness, $z=1.48$, $p<0.05$; for height, $z=2.08$, $p<0.05$).

Given a student's set of strategies in the first trial, the number of matching strategies in the second trial was found, and a proportion of matching strategies calculated. For the purposes of a chi-squared test, those students with a proportion of matching strategies equal to or greater than 0.75 were labelled 'low variability' while those with a proportion less than 0.75 were labelled 'high variability'. A chi-squared test with variables 'gender, m/f' and 'variability, high/low' (see Table 14) showed that more boys showed low variability and more girls showed high variability ($\chi^2=4.07$, $p<0.05$).

Table 14: Distribution of boys and girls across two levels of variability

	High Variability	Low Variability
Girls	15	9
Boys	4	10

6.3.3 Effects of achievement level, gender and order of contexts on test scores

In order to isolate gender as an independent factor in the relationship with pattern of strategy use, the proportion of matching strategies across trials was used as a dependent variable in regression tests with independent variables including Key Stage 2 mathematics SAT score $F(1,47)=0.124$, $p=0.729$, Year 9 Key Stage 3 mathematics SAT target $F(1,47)=0.031$, $p=0.862$, average time spent on task in 1^{st} trial $F(1,47)=2.147$, $p=0.152$ and in 2^{nd} trial $F(1,47)=1.267$, $p=0.268$ and school $F(1,47)=0.895$, $p=0.350$.

A comparison between the numbers of strategies used by boys and girls was made, in order to confirm that lower variability of strategy use by boys was not due to use of fewer unique strategies altogether (see Table 15 for means and standard deviations). There were no differences in the number of different strategies used by boys and girls in either the first trial ($df=47$, $t=2.141$, $p=.202$) or in the second trial ($df=47$, $t=.071$, $p=.855$).

Table 15: Means and Standard Deviations (in italics) of the numbers of unique strategies used by boys and girls in each trial

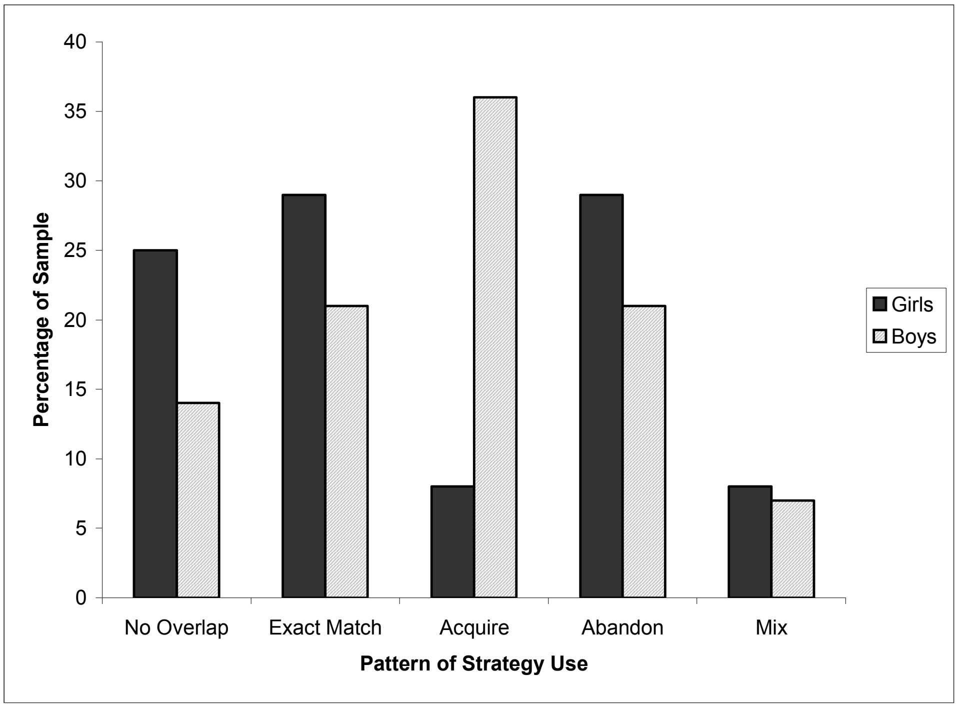
	Trial 1	**Trial 2**
Boys	1.4 *0.5*	1.8 *0.6*
Girls	1.8 *0.6*	1.4 *0.8*

6.3.4 Patterns of strategy use

An analysis of the distribution of patterns of strategy use helps to identify some of the ways in which boys and girls differ in their development of thinking about the problem set (see Figure 6). It is not possible to conduct a chi-squared test on this distribution, as the expected frequencies are too small – however the difference between the distributions observed here and those observed in study 2 are striking. The 'acquire' category seems to be the only category in which a difference between boys and girls can be seen. In all other categories the percentage of boys and girls who fall into any category is approximately equal once the category 'acquire' is removed.

The differences in the results of this study compared with studies 1 and 2 are such that the think-aloud protocols produced in the present study cannot be thought of as providing further information regarding the effects found in the previous studies. In fact, any findings resulting from analysis of the think-aloud protocols are overshadowed somewhat by the effect of the inclusion of think-alouds in this study on the gender effect under investigation. However, some comments can be made regarding children's answers to questions given as part of the interview.

Figure 6: Graph to show the distribution across categories of patterns of strategy use by gender



6.3.5 Time on task

Time on task was compared across gender. Means and standard deviations are given in Table 16.

Table 16: Means and standard deviations (in italics) of average time on task in seconds during trial 1 and trial 2 by gender

	Trial 1	**Trial 2**
Boys	31.5 *11.5*	20.5 *7.5*
Girls	33.1 *20.7*	21.7 *9.7*

There were no significant differences between the times taken to solve problems in either trial between boys and girls (Trial 1: F(1,47)=.071, p=.792; Trial 2: F(1,47)=.167, p=.685).

Time on task was also compared across levels of strategy variability. A separate test was used, as gender and strategy variability are known not to be independent. As in section 6.3.2, participants were considered to have shown a low level of variability if the proportion of matching strategies across trial was greater than or equal to 0.75. Means and standard deviations are given in Table 17.

Table 17: Means and standard deviations (in italics) of average time on task in seconds during trial 1 and trial 2 by level of variability

	Trial 1	**Trial 2**
High Variability	28.4 *11.6*	18.9 *7.4*
Low Variability	36.5 *21.8*	23.6 *9.8*

There were no significant differences between the times taken to solve problems in either trial between participants showing high and low variability (Trial 1: F(1,47)=2.046, p=0.161; Trial 2: F(1,47)=2.765, p=0.105).

6.3.6 Interview data

The interviews were intended to generate additional data to help explain the relationship between gender and variability.

One aim of the interviews was to establish whether differences in strategy variability are associated with the memory that children had at the time of the second trial for the questions set at the time of the first trial. To this end, after they had answered all the questions in the second trial, all children were asked to say all that they could about any similarities or differences that they could think of between the problem sets encountered in the two trials. All children remembered that the previous week's questions had involved similar questions, specifically that the questions had involved making judgments about the relative rates of increase of two lines, both rising. A majority of the children remembered that the contexts of the problem set were different (61% overall, 58% of girls, 63% of boys). There was no association between children's memory for differences between the problem sets and children's level of strategy variability (df=1, χ^2= 0.45, p=0.5, see Table 18).

Table 18: Distribution of children across variability categories according to memory for differences in context between problem sets

	Remembered Differences in Context	**Did Not Remember Differences in Context**
High Variability	11	8
Low Variability	13	6

When questioned about their answers, specifically about whether their strategies had changed between trials, all children said that the way that they had answered the questions was the same for trial 2 as for trial 1. It is not possible to tell, for those children who used a different set of strategies in trial 2 to that used in trial 1, whether that was due to poor memory for strategies used in trial 1, or to a false belief that strategies were the same.

A third aim of the interview was to establish whether children understood the contradictions arising as a result of the application of competing strategies to isomorphic questions in the first and second trials, and determine whether an understanding of such contradictions was associated with children's strategy variability. The last part of the interview involved showing both problem sets together to the student, and explaining the isomorphic nature of the problem sets.

On seeing both problem sets together, children were all able to understand that the only difference between them was the context. Where children's answers to corresponding questions were different, they were asked if they still agreed with both answers. The vast majority of children responded with a desire to change one or other answer so that they were the same. 86% of children said that they would change all answers where there was a difference, in order that answers to corresponding questions were identical. Of these children, 80% preferred the answer given in the second trial.

Those children who did not feel it necessary for corresponding answers to match all gave similar reasons for leaving their answers as they were. These children believed that the differences in the wording of the question, due to the differences in context between problem sets, were such that they were asking different things. This was despite making efforts to ensure that children understood that the problem sets were devised such that they were asking exactly the same question. The most common problem that children had in this area was in interpreting the phrases 'up to…' and 'before…' as referring to identical regions of the graphs in trials 1 and 2. The children who were happy for corresponding answers not to match argued that when the question was phrased ‘up to point N’, then that meant they should be looking at point N itself rather than the region of the graph to left of point N and that when the question was phrased ‘before 1984’, that the question referred to the region of the graph to the left of 1984 rather than the point of the graph *at* 1984. These instances of children’s seeming acceptance of contradiction may then be thought of as a problem of interpretation of language rather than mathematical objects. While this misinterpretation of the language used for the questions might seem to cause a problem for the analysis of differences in strategy use, it is important to note that this affected only 5 children, who showed a contradiction in only one question each. This does not affect the validity of the analysis earlier in this section. Of those five children, three were girls and two boys. This would seem to negate the possibility that an ability to perceive contradiction in multiple strategy sets might act as a mediator of the association between gender and strategy variability.

6.3.7 Summary

The results presented in this section show an effect of gender on strategy variability. There is evidence for an interesting reactivity to the use of think-aloud protocols that appears to have reversed the effect observed in studies 1 and 2. In the present study, boys have shown lower levels of variability than have girls.

Patterns of strategy use appear to be different to those observed in studies 1 and 2. The only difference between boys and girls in this study in terms of pattern of strategy use was that a substantial number of boys (and many more boys than girls) added one or more new strategies to their repertoire between sessions. Unlike in previous studies, few children both acquired *and* abandoned strategies between sessions.

The results show no mediating effect of time on task, memory for problem set or children's perceptions of contradiction on the association between gender and strategy variability. The results therefore add to the body of evidence suggesting a genuine effect of gender on strategy variability, although they indicate that the relationship is likely to be complex, given the reactivity to think-alouds.

6.4 Discussion

It is clear from the results that the relationship between gender and strategy variability is not as simple as it appeared to be on the basis of the first two studies reported in this thesis. It is clear that the think-alouds, although generating very detailed data on children's strategy decisions, are not helping answer the intended questions. While there is good evidence again for a gender effect on levels of variability over time, the effect appears to be in the reverse direction, with girls showing higher levels of strategy variability compared with boys. What is especially interesting about the results of this study is that there is a complete crossover with regard to the observed effects. It is not that surprising that think-aloud protocols have had an effect on strategy use, but what is surprising is that they have had opposite effects for boys and girls. This crossover effect will be discussed in greater detail below in section 6.4.2.

6.4.1 Strategy Variability and Gender

Before further discussion of the crossover effect, it is important to focus on the association between gender and strategy variability. The results again show gender to be a strong predictor of strategy variability. Again, gender has been shown to affect variability independent of children's level of success in answering questions in the problem set and independent of the suitability of strategy selected. The gender effect was shown to be the same in each of the four classes of children tested, in two schools.

The primary aim of this study was to assess the role of various potential co-variants with gender in explaining the gender/strategy variability relationship. These included memory for both problem set and strategy set, ability to perceive contradictions in strategy sets and time on task. Each of the potential co-variants will be considered in turn.

Individual differences in memory were though to have the potential to explain differences in strategy variability for two main reasons. Firstly, a clearer memory for either the problem set encountered or the strategy set employed in the first trial was thought to be a potentially good predictor for variability. This was because it seems that without a good reason to change, the default behaviour expected from children is likely to be to use the same set of strategies in the second trial as in the first. It is more likely that children will use the same set of strategies in the second trial as the first if they are able to remember both the problem set encountered and the strategy set employed in the first trial when they come to solve the problem set administered in the second trial.

In order to investigate any potential mediating effect of memory on the relationship between gender and variability, children were asked questions during the interview that related to their memory for both the problem set and strategy set from the first trial. There was no evidence that memory for either the problem set or the strategy set co-varied with gender. Nor was there any evidence that memory for either the problem set or the strategy set was a predictor of strategy variability. There was sufficient variance in children's memory for the problem set encountered in the first trial that any predictive power would likely have been identified had it been there to find.

There was very little variance in children's memory for the strategy set used in the first trial. Children almost exclusively believed that the strategy set used in the first trial was identical with that used in the second trial when questioned. Of course, for those children who actually did use the same set of strategies in both trials, this assessment was correct. Therefore it remains possible that those children who used the same set of strategies in both trials were the children who could accurately remember the set of strategies used in the first trial. It is not possible to come to any definite conclusion regarding any effect of children's memory for the strategy set without an improved measure of that memory. On balance, from the evidence obtained in this study, there is certainly no evidence for any such effect. The most reasonable conclusion on the basis of the evidence presented here is that children are equally poor at remembering the strategy set used in trial 1, and that it has little, if any, effect on the strategy set used in trial 2.

Just as there has been no evidence to suggest an effect of memory on strategy variability, it has been possible to rule out children's perceptions of contradiction as a possible predictor of differences in strategy variability in this study. On the basis of previous studies reported in this thesis, it was thought that there may be a relationship between children's understanding of the incompatibility of some strategy sets and their level of strategy variability. It was predicted that girls were more likely to be able to perceive contradiction or incompatibility in the strategy sets used to solve the two sets of problems, thereby explaining their lower level of variability in previous studies. In this case, the only evidence in question was the existence or not of any relationship between perception of contradiction and gender. It was not possible to assess any relationship between perception of contradiction and strategy variability, as all those participants exhibiting a lack of perception of contradiction necessarily employed a different strategy set in trial 2 compared with trial 1.

The vast majority of children were very quick to decide that the two sets of answers should be the same, due to the isomorphism of the two problem sets. Where children were shown their two sets of answers, they were almost all able to decide that the two sets of answers should be the same. Only five children believed that an answer should be different between the two sets of problems. Making up these five were two boys and three girls. In each case, this was due to their belief that the questions referred to different regions of the graph. With so few children exhibiting a lack of perception of contradiction amongst strategy sets, and those that did being equally distributed across gender, there was no evidence for any mediating effect of perception of contradiction on the association between gender and strategy variability.

Children's temperament and motivation were also thought to have the potential to represent a mediating factor in the relationship between gender and variability. There is evidence to suggest that boys and girls differ in their attitudes towards mathematics, and also that some differences in strategy use can be explained by differences in temperament. In order to assess the potential of these aspects of children's behaviour to help explain the relationship between gender and variability, the time taken by children to answer each question was recorded.

Any association between time on task and either gender or variability would have provided evidence for some mechanism responsible for the relationship between gender and variability. If present, it would not have determined the precise nature of the mechanism involved, whether it was children's level of inhibition, impulsiveness or motivation, for example that was responsible for differences in strategy variability. Its absence, on the other hand, would provide evidence that no such mechanism was present.

Time on task varied systematically across trials, as might be expected. Problems in trial 2 were generally solved quicker than those in trial 1. This effect is probably due to a combination of things, including familiarity with both the experimental environment and the nature of the problem set administered.

It was possible to test for associations between time on task and both gender and level of variability. There was a high degree of variance amongst children's average times taken to solve the problems. Therefore it would have been reasonable to expect to detect any association of time on task with either gender or strategy variability should such an association be present. Analysis of time on task was thorough; in making a judgement as to the role of time on task in the gender effect, analyses of average time taken on problems in trial 1, in trial 2 and overall were conducted, as were analyses of the distribution of times taken to solve problems in each set. The results show no relationship between time on task with gender and no relationship between time on task and strategy variability. This is a fairly substantial finding, as it would seem to rule out a number of potential explanations of the association between gender and strategy variability.

6.4.2 Reactivity to Think-alouds

Comparing the results of this study with those of studies 1 and 2, the boys have become less variable while the girls have become more variable. The extent of the crossover effect can be seen by looking back at the correlations calculated between the uses of steepness and height in the first and second trials, both in the present study and in study 2. In each study, Fisher transformations show that there is a gender difference in variability. If we compare studies, for both boys and girls there is a significant difference in correlations between uses of both height and steepness in the first and second trials. These observations are summarised in Table 19.

Table 19: Summary of gender effects on strategy variability in studies 2 and 3. Each high-low variability pairing shows significantly different level of variability

	No think-alouds	**Think-alouds**
Girls	Low variability	High variability
Boys	High variability	Low variability

It is difficult to speculate as to the possible mechanism by which the crossover effect arises. First of all it is important to decide how much confidence can be placed in the effect itself. The best way to think of this crossover effect is by combining the data accumulated in the present and previous studies, and thinking of the two studies as one between groups study, with two independent variables, gender and use of intervention (intervention being the use of think-aloud protocols and retrospective reports) , each with two conditions. Viewed as one study, it is clear that the main conclusions of the study would be based on a significant interaction between gender and think-aloud intervention. It seems that the two studies between them provide enough evidence for that interaction.

One of the main difficulties in determining the mechanism by which the crossover effect arises, is the number of variables that co-vary with the use of think-alouds. When think-alouds are used, children spend more time working on each problem, they spend more time thinking about each problem after having given an answer, they are under more pressure to perform and there is a greater load on working memory. There are surely many other factors that are associated with the use of think-aloud protocols.

So one possible mechanism for the crossover effect could involve math-anxiety, mentioned in the introduction to this study. Girls' higher level of math anxiety compared with boys' suggests a possible cause of differences in strategy variability. The arguments given in the introduction concerning memory, where girls' better memory for events was argued to be associated with a higher likelihood of reusing a strategy set, assume that the strategies used by children during the first trial resulted in a positive outcome. However, as the experimenter gave no feedback regarding the correctness of answers to the children, it is not clear that children necessarily believed that a positive outcome had been achieved at the end of the first trial. Casey, Nuttall and Pezaris (2001) showed that for mathematics tasks where a gender difference was present, self-confidence in mathematical ability accounted for a significant amount of that difference. Therefore it is possible that with the increased analysis of questions and strategies associated with the think-alouds and retrospective reports, girls were more likely to doubt their answers in the first trial and therefore use a new set of strategies in the second trial. Conversely, boys, with high levels of self-confidence in mathematics, would be more likely to believe in the correctness of their answers in the first trial and thus use a similar set of strategies in the second trial. It is easy to see that this is just one possible explanation of many, and a great deal of further work will be required to unravel the effects observed here.

It is worth spending a bit more time discussing the effect of think-alouds, however, as it is usually reported that think-alouds have no effect on the problem-solving behaviour of participants. This is usually shown to be the case by comparing samples of participants either thinking aloud or not thinking-aloud, and demonstrating that distributions of strategy choices are similar for both samples. It can be shown for the present and previous studies that the distributions of strategies are the same for the two studies, with and without think-alouds. The effects of the think-alouds only become apparent when the pattern of strategy use over time is analysed. This may be a crucial finding for future research. If think-alouds *are* responsible to any degree for determining problem-solving behaviour in subsequent trial, this calls into question the use of think-alouds in any microgenetic or longitudinal study investigating problem solving behaviour, where there are not suitable controls in place to account for the use of such protocols. It is not sufficient to compare the distributions of strategy choices in two samples and conclude, on the basis of similar distributions, that think-aloud protocols have had no effect on behaviour.

6.4.3 Conclusion

In summary, on the basis of the present study, three conclusions can be drawn. Firstly, it is becoming clearer that there is an association between gender and strategy variability. Secondly, the mechanism behind that association is complex, and involves some interaction with situational factors. Thirdly, think-alouds have an effect on strategy choices, differentially according to gender.

This study has certainly been able to confirm the fact that there is an interesting relationship between gender and strategy variability and has been able to focus the search for mediating factors, co-varying with gender. If gender were an independent cause for differences in strategy variability, then the effect would likely be consistent despite minor changes to the procedure of experiments. This study has shown that there is likely to be an interesting mechanism lying behind the apparent relationship between gender and variability. At the least, it is possible to say that there is some interaction between gender and questioning environment that affects strategy choice behaviour.

Although the think-aloud protocols are not telling us about the association that we had hoped they would, their analysis has still proved to be enlightening. It appears that memory for either the problem set or the set of strategies used in trial 1 does not have an impact on strategy variability. All children said that they had used the same set of strategies in the second trial as in the first. Also there was no difference between boys' and girls' memory for the similarities and differences between the two problem sets.

The next stage in this investigation will be designed in order to add confidence to the findings presented here. The more unexpected that a set of results is, the more necessary it is to replicate them. The crossover effect was certainly unexpected – it will be necessary to test this effect for robustness. The most obvious way do this in the first instance will be to repeat the same procedure, for the same children, after a period of time has passed.

The gender effect on strategy variability without think-alouds, observed in study 1, was replicated in study 2. Similar findings in study 2 provided confidence in and elaboration of the findings reported in study 1. The replication of the opposite gender effect found here in study 3 provides a number of opportunities to enhance an emerging understanding of the association between gender and strategy variability. Most importantly, it will both show whether the effect of think-alouds continues to have the same effect on the gender-strategy variability relationship and show whether individual differences in strategy variability are durable (i.e. whether children show similar levels of strategy variability in separate studies). As few changes as possible must be made to the procedure reported here, so that any changes in the results can be explained accurately.

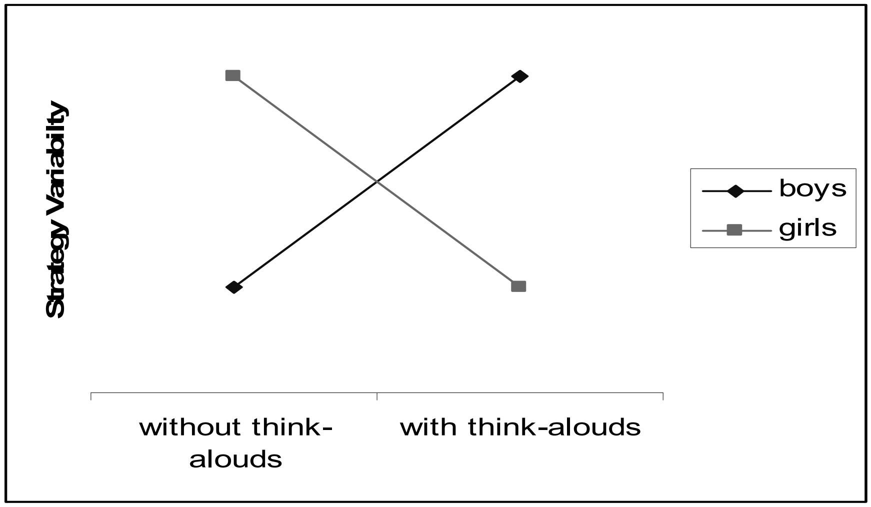
7 Study 4

7.1 Introduction

The previous study showed an interesting effect that requires some clarification. The introduction of think-aloud protocols to the data collection procedure had the effect of reversing the gender effect on strategy variability. In studies 1 and 2, in which think-aloud protocols were not used, boys showed a high level of strategy variability compared with girls. In study 3, which differed procedurally only with the introduction of think-alouds, girls showed a high level of variability compared with boys.

The scale of this cross-over effect was large enough to suggest a real effect of reactivity to the think-alouds. Analysis of the differences in variability of boys and girls with and without think-alouds showed significant differences between all groups, as in Figure 7.
The nature of this reactivity effect is surprising due to the different direction of effect observed for boys and girls. If both boys and girls were affected by the introduction of think-alouds in the same way, it might be easier to understand and explain the effects. The fact that boys and girls appear to be affected in opposite ways by the introduction of think-alouds suggests that some interesting mechanisms are at work.

Figure 7 - strategy variability of boys and girls with and without think-aloud protocols



Despite the size of the observed reactivity effect, it is difficult to accept it at face value, considering the weight of evidence for the validity of think-aloud protocols. Ericsson and Simon (1993) claim that there is no reactivity to think-aloud protocols, however, research discussed earlier (see section 5.1) suggests that verbal protocols may indeed have some unwanted effects on behaviour. This is a very important methodological issue as verbal protocols are often used to investigate children's mathematical thinking. If these verbal protocols are altering the behaviour that is being measured, then serious issues of validity are raised.

It will be necessary then to increase the weight of evidence suggesting reactivity. A key aim of the present study will be to generate more confidence in the crossover effect whereby when think-aloud protocols are employed, girls show a high level of strategy variability compared with boys.

A second issue that requires investigation is the durability of the effects observed so far. The question of whether or not strategy variability is a consistent factor in children is vital in order to establish the usefulness of this association. If children's level of strategy variability were inconsistent, it would be difficult to make practical use of any knowledge about individual differences in strategy variability. If, on the other hand, children's level of strategy variability can be shown to be fairly consistent, it will be possible to imagine some possible practical applications of the findings described so far for the classroom. It would be possible, for example, to diagnose individual children's typical level of strategy variability and structure those children's interventions and support accordingly. Of course, this is a little way off at present, with a great deal of further research required before any findings presented in this thesis can be applied. A second aim of the present study then will be to establish to what extent children's level of strategy variability is invariant over time.

In order to satisfy both aims, this study will take the form of a replication of study 3, using the same problem sets and the same children. The only difference between the two studies will be the fact that the present study was conducted six months later. The hypothesis for this study will be that results for this study will match those of study 3. In other words, it will be predicted that children's levels of strategy variability in response to the problem sets under investigation remain stable over time.

7.2 Method

7.2.1 Design

The independent variable in this study is gender. The dependent variable is strategy variability. The main prediction is that there will be an effect of gender on strategy variability, with girls showing greater variability than boys.

There is a second aspect to the analysis of the results in this study, involving a comparison of students' levels of strategy variability in the present study with that in the previous study, conducted six months earlier. The hypothesis for this aspect of the study is that there will be a positive correlation between children's levels of strategy variability in study 3 and in the present study.

7.2.2 Participants

The participants in this study were the same as in the previous study. There were some children who had either left the school or were absent during the time in which testing took place. Therefore participants were 17 girls and 11 boys aged 13 or 14 from two schools in Nottinghamshire.

7.2.3 Tasks and measures

The tasks used in this study were the same as those used in study 3 (see section 6.2.3). There were two sets of six problems, one set involving a sparse, abstract context, the other involving a more realistic context. The problems all involved children making judgements about rates of change on the basis of straight line graphs.

To ensure that context order effects were accounted for, each context order group (abstract first or realistic first) from the previous study was split in half at random. One half of each group saw the abstract set first in the present study, while the other half of each group saw the realistic set first.

Strategy variability was determined by comparing the strategies used in the first and second sessions. The proportion of matching strategies was found for each participant. Each strategy in each session could be part of only one matching pair.

7.2.4 Procedure

As one of the major aims of this study was to compare participants' levels of strategy variability with those measured in the previous study six months earlier, the procedure in this study was identical with that of study 3. Therefore, participants were all asked to think aloud while working on each problem, and upon completion of the problem set were asked for a retrospective report on the strategies used to solve each problem. All sessions were videotaped. For further details, refer to section 6.2.5.

7.3 Results

7.3.1 Distribution of strategies used

Table 20 shows the frequencies of strategies used to answer the problem sets in this study, according to session and gender.

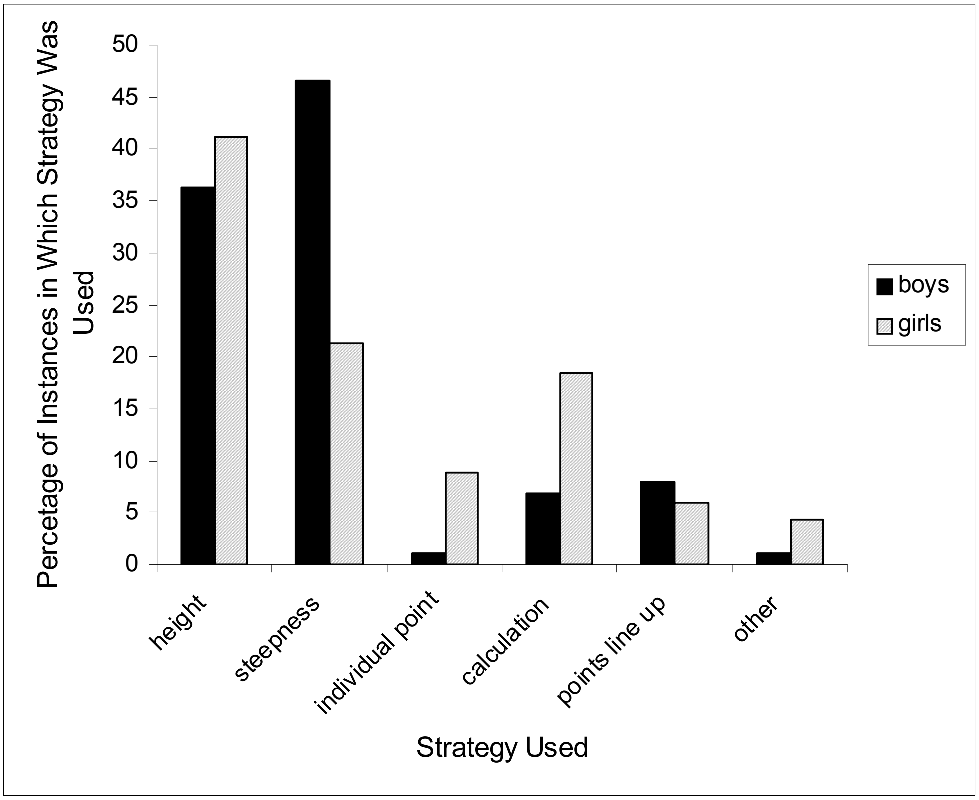
Table 20: Frequencies of strategies used according to session and gender (study 4)

		Height	Steepness	Individual Point	Calculation	Points Line Up	Other	Total
Session 1	Boys	16	21	0	4	3	0	44
	Girls	26	17	6	11	7	1	68
Session 2	Boys	16	20	1	2	4	1	44
	Girls	30	12	6	14	1	5	68

The distribution of participants across strategies was similar for each session (χ^2=8.55, df=5, p>0.05). There were differences, however, in the distribution of strategies used by boys and girls (χ^2=21.96, df=4, p<0.001) – for this test, the 'other' category was eliminated due to expected frequencies lower than 5. Therefore, Figure 8 shows the distribution of strategies used, collapsed across session.

Figure 8 shows that the majority of strategies used involved the relative heights or gradients of the two lines. This is very similar to the distribution found in previous studies reported in this thesis (see for example, Figure 5). The principal differences between the boys and girls are that more boys than girls use a strategy involving the relative steepness of the lines, while more girls than boys use strategies involving calculation, and the fact that the lines meet at a point.

Figure 8: Graph to show distribution of participants across strategies, by gender, collapsed over session (study 4)



It is clear from this data that there was little to no convergence on more appropriate uses of strategy during the six months since study 3 was conducted.

7.3.2 Strategy variability

Analyses conducted in study 3 were repeated. This involved the calculation of correlations for the use of the two most commonly used strategies across the two sessions. Table 21 and Table 22 show correlation coefficients for boys' and girls' use of relative steepness and relative height across sessions.

Table 21: Correlation between use of steepness in 1st and 2nd session (study 4)

	Boys	Girls
Pearson's correlation coefficient	0.988	0.752
Significance	0.000	0.001

Table 22: Correlation between use of height in 1st and 2nd session (study 4)

	Boys	Girls
Pearson's correlation coefficient	1	0.653
Significance	0.000	0.004

It appears that boys are more likely than girls to use a strategy in the second session given its use in the first session. Fisher's transformation shows that for both strategies, correlations are significantly stronger for boys than for girls (height: $z=6.81$, $p<0.001$; steepness: $z=3.56$, $p<0.001$). It should be pointed out, however, that the correlations shown for girls are still very strong.

Again as in study 3, a comparison was made of the distributions of boys and girls across levels of variability. Table 23 shows this distribution.

Table 23: Distribution of boys and girls across two levels of variability (study 4)

	High Variability	Low Variability
Girls	12	5
Boys	3	8

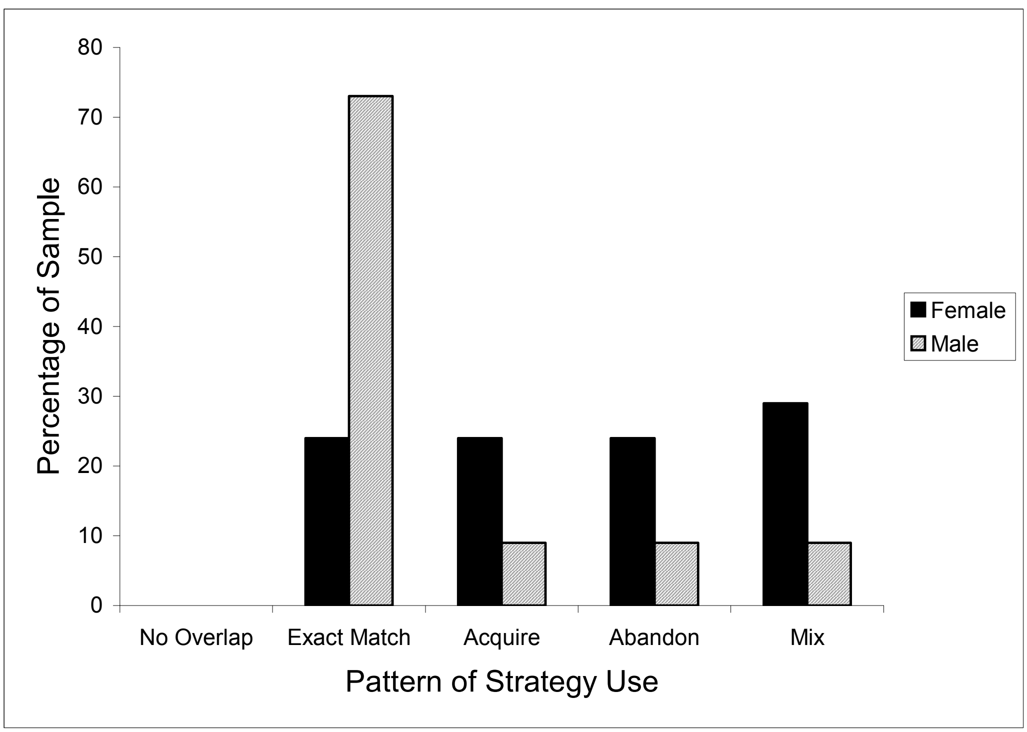
A chi-squared test showed that more boys were in the 'low variability' group and more girls were in the 'high variability' group ($\chi^2=5.04$, df=1, $p<0.05$).

As demonstrated above, this reduction in variability is not due to students' selections of more suitable strategies.

7.3.3 Patterns of strategy use

Figure 9 shows the distribution of patterns of strategy use across participants. It is apparent that the boys in this study were much more likely to use the exact same set of strategies in the second trial as in the first than were girls. A second interesting feature of this distribution is that no children now fall into the 'no overlap' category. There are no children that use a completely different set of strategies in the second trial than in the first in this study.

Figure 9: Graph to show the distribution across categories of patterns of strategy use by gender (study 4)



7.3.4 Comparisons with study 3

The main objective of this study was to determine whether strategy variability was a durable characteristic in children's problem solving behaviour. In order to test the durability of the gender effect on strategy variability, Pearson's correlation coefficient was calculated for measures of strategy variability for participants' problem solving in Study 3 and in the present study. The strategy variability score was obtained by counting the number of strategies in trial 1 that correspond with strategies used again in trial 2 and therefore could be between 0 and 4. There was a positive correlation between the number of matching strategies in study 3 and in study 4 ($r=0.336$, $p=0.066$).

Comparison of correlations of the use of relative steepness and relative height across sessions suggests that strategy use is becoming less variable for both boys and girls. This can be seen in Table 24. The change is large for both boys and girls, but should perhaps be considered to be more dramatic for the boys as they show almost zero variability in study 4.

Table 24: Comparison of correlations of strategy use across sessions in study 3 and 4

		Study 3	Study 4
Boys	Steepness	0.623	1
	Height	0.713	0.988
Girls	Steepness	0.176	0.653
	Height	0.120	0.752

Comparisons of patterns of strategy use between study 3 and 4 also show some notable differences (see Table 25).

Table 25: Distribution of patterns of strategy use for boys and girls (percentages) for the present study and for study 3 (in parentheses)

	Boys	Girls
Exact match	73 (21)	24 (29)
No overlap	0 (14)	0 (25)
Acquire	9 (36)	24 (8)
Abandon	9 (21)	24 (29)
Mix	9 (7)	29 (8)

It appears that these data reinforce one of the conclusions drawn above, that boys' variability was lower than girls' in study 3, and has reduced to a greater extent in the six months between that and the present study. The distribution of boys across categories is much more strongly skewed towards the 'exact match' category than that of girls, who are fairly evenly distributed across all categories bar 'no overlap'.

7.3.5 Summary

The key results of this study concern the comparisons of data with that collected in study 3. The data suggest that strategy variability can be considered a durable characteristic of children's problem solving behaviour.

There is also further evidence here to suggest that strategy variability can be considered independent of the correctness or appropriateness of the strategy selected. Over six months, there has been a general decrease in variability, while the distribution of strategies used has remained stable across the sample.

7.4 Discussion

The aim of this study was to investigate the durability of strategy variability as a characteristic of children's thinking. The question of durability is important as any answer will guide the development of future research in this area. If the association between gender and strategy variability described in the preceding three study reports is to be practically useful in the classroom, then it is important to know to what extent strategy variability is a stable characteristic in children's mathematical thinking.

In this study, the same children as in study 3 were retested using the same problem materials, six months later. Analysis of the results has helped to show that children's level of strategy variability, at least for this age group in response to the problem set under investigation, is a fairly stable characteristic.

An analysis of the correlation between levels of strategy variability observed in study 3 and in the present study showed that a child's level of strategy variability is well predicted by their level of strategy variability six months previously. Again, as in each of the previously reported studies in this thesis, the reported group differences in children's level of strategy variability were independent of the strategies selected, children's level of success in answering the problems set, the school and class to which children belonged and children's level of achievement in school mathematics.

Analysis of changes in children's patterns of strategy use between study 3 and the present study show some interesting trends. What is most interesting is that the trends in patterns of strategy use between study 3 and the present study appear to be different for boys than for girls. The boys seem to have become much less variable over the six months than have girls. The distribution of boys across categories of patterns of strategy use has changed from one in which there were roughly equal numbers in each category to one in which they are predominantly in the ‘exact match’ category.

8 Study 5

8.1 Introduction

This study was conducted in order to ascertain the effect, if any, that ability has on the relationship between gender and variability for the problem set in question. In each of the studies conducted as part of this thesis so far, children of similar abilities have participated. All children taking part in each of the studies 1-4 were in Year 9 (13-14 years old) and expected to achieve a level 5 in the Year 9 Mathematics SAT. It has been possible to say with increasing certainty that for these children, gender is a predictor of strategy variability, although the exact nature of the relationship will need increased clarification in the future.

It is not yet clear whether there is an effect of ability on the relationship between gender and variability. The hypothesis for this study is that the observed effect of gender on variability will be restricted to a particular ability range, representing the zone of proximal development or ZPD (Vygotsky, 1978).

The zone of proximal development can be identified with the set of problems that cannot yet be solved by a child independently, but can potentially be solved with guidance either from an adult or a more capable peer. In the literature review earlier in this thesis, the rate of change problem set used for each of studies up to now was described as suitable due to, among other reasons, the fact that the problems were accessible, but not trivial, for the participants involved. The range of problems that fit this description, accessible but not trivial, corresponds very closely to Vygotsky's description of problems that fall within the ZPD. Those problems that are accessible but not trivial are exactly those problems that are solvable by children with guidance. Participants for the previous four studies have been carefully selected to ensure that they understand graphs, evidenced by their ability to give a y-value in response to an x-value and vice versa, also their ability to calculate incremental increases in y-value when given a pair of x-values. Care has also been taken to ensure that participants have not been introduced to concepts such as gradient, rate of change or compound measures in the classroom. What should we expect to find if a similar study was conducted, without that same careful selection process?

There is increasing evidence that the process of cognitive change is associated with an expansion followed by a contraction of the strategy repertoire (Siegler & Shipley, 1995). The children selected to participate in the previous studies reported in this thesis were selected to be at the point of cognitive change. The present study will address the question of the patterns of strategy use shown by children with abilities higher and lower than that range investigated up to now.

Children who fall below the ability range addressed up to now should be thought of as having reduced access to the problem set. Children who are expected to achieve below a level 5 in the Year 9 Mathematics SAT are likely to have difficulty interpreting graphs. Pilot studies described earlier in this thesis showed that children expected to achieve below level 5 were less likely to be able to accurately return a y-value in response to an x-value and struggled to calculate incremental increases in y-value between a pair of x-values. This reduced access to the problem set should be expected to lead to a reduced set of strategies, compared to the samples tested up to now, likely to lead to generally incorrect answers. Children with lower levels of ability should be more likely to be in the ‘acquire’ or ‘no overlap’ categories of pattern of strategy use.

Children with a level of ability above that addressed up to now should be expected to be moving toward the point at which the problem set becomes trivial. Children with higher levels of ability, who are more advanced mathematically, should be predicted to be more likely to be found in either the ‘abandon’ or ‘exact match’ category of pattern of strategy use.

What of the effect of varying ability on the effect of gender on variability? It seems likely that the gender effect will be reduced for those students for whom the problem set in question does not fall within the ZPD. Little to no gender effect is predicted for those children with a lower level ability than previously tested, with limited access to the problem set. For children with a higher level of ability, the gender effect will also be predicted to be small, if one exists, as many more children in this group will be generating correct answers, and therefore demonstrating a reduced repertoire of strategies in comparison to those children who participated in studies 1-4.

In summary, the gender effect is predicted to exist only in samples of children for whom the problem set is within the ZPD. This study aims to show that children for whom the problem set is outside the ZPD have differing distributions of patterns of strategy use, depending on whether their ability range is high or low, and therefore will show a reduced gender effect in comparison with samples of children tested previously.

8.2 Method

8.2.1 Design

The independent variables in this study were gender and level of achievement in mathematics. The dependent variables were strategies used and strategy variability. The main aim of the study was to find out to what extent gender had an effect on strategy variability in each achievement group. It was predicted that, contrary to findings with a sample with an intermediate level of achievement, both the high and low achievement groups would show only a limited effect of gender on variability.

A second aim of the study was to compare the distribution of students across strategies for both achievement groups and analyse differences in distribution that could help to describe development of understanding in this domain.

8.2.2 Participants

Participants in this study were 85 students, aged 13-14 years old, from 4 classes in a school in Nottinghamshire. These participants formed four groups, distinguished by gender and by level of achievement. The two levels of achievement of the students were chosen to create groups either side of the children participating in each of the studies reported in this thesis up to now

(expected to achieve a level 5 in the Year 9 Mathematics SAT). Therefore, the high achievement group consisted of children expected by their teachers to receive a level 6 or higher in their Year 9 Mathematics SAT. The low achievement group consisted of children expected to receive a level 4.
The high achievement group consisted of 32 girls and 20 boys, while the low achievement group consisted of 19 girls and 12 boys.

8.2.3 Tasks and Measures

The tasks used in this study were identical with those used in previous studies reported here (for example, see 6.2.3). There were two sets of six problems, one set involving a sparse, abstract context, the other involving a more realistic context. The problems all involved children making judgements about rates of change on the basis of straight line graphs.

To ensure that context order effects were accounted for, each context order group (abstract first or realistic first) from the previous study was split in half at random. One half of each group saw the abstract set first in the present study, while the other half of each group saw the realistic set first.

Strategy variability was determined by comparing the strategies used in the first and second sessions. The proportion of matching strategies was found for each participant. Each strategy in each session could be part of only one matching pair.

8.2.4 Procedure

This study was intended to generate results that could be compared to those found in the previous studies reported here, especially those in study 1. Therefore the procedure was designed to be as similar as possible to that used in study 1 (see section 3.2.4).

Half of the participants in each group were administered the abstract problem set first. The other half of the participants were administered the realistic problem set first. The problem sets were completed in students' usual mathematics classes under examination conditions; participants were not able to communicate with each other, or to see other students' answers. Participants were allowed ten minutes to complete all six questions. For each question, participants were instructed to provide both an answer and an explanation for that answer. Participants were informed that the explanation for the answer was very important and that they should make sure to complete this section as completely as possible.

Whichever problem set was not completed by a participant in the first session was administered to that participant the following week. Again, the problem sets were completed in students' usual mathematics class, under examination conditions.

8.3 Results

8.3.1 Distributions of strategies used

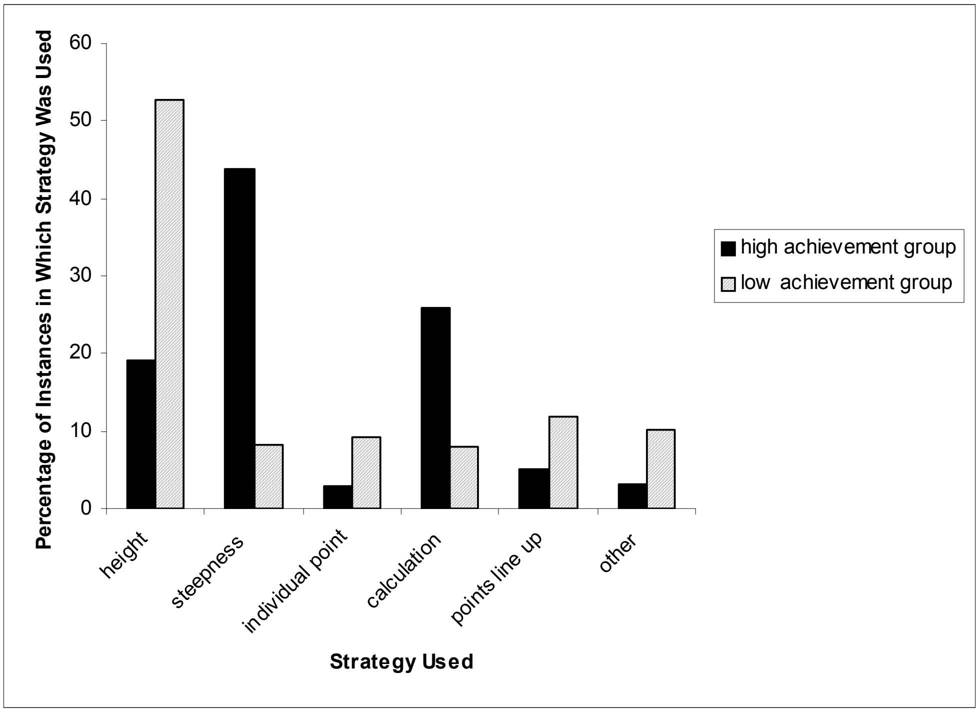
Table 26 shows the frequencies of strategies used by children in response to the problem sets.

Table 26: Frequencies of strategies used by group and session (study 5)

	Height	Steepness	Individual Point	Calculation	Points Line Up	Other
High Achievement – Session 1	40	94	12	50	21	13
High Achievement – Session 2	40	88	5	58	11	6
Low Achievement – Session 1	64	22	24	21	31	27
Low Achievement – Session 2	75	11	13	8	13	12

There are no differences in the distributions of strategies used in each session either in the high achievement group (χ^2=8.29, df=5, p>0.05) or in the low achievement group (χ^2=8.59, df=5, p>0.05). There are, however, significant differences between the distributions of strategies used by the high achievement and the low achievement group (χ^2=166.84 , df=5, p<0.0001). Figure 10 shows the distribution of strategies, according to achievement group.

Figure 10: Distribution of strategies across instances of strategy use, according to achievement group (study 5)



The vast majority of strategies used by the low achievement group involve the relative heights of lines. For the high achievement group, on the other hand, the most common strategies used involve the relative steepness of the lines, followed by the use of calculation.

8.3.2 Strategy variability

Analysis of gender effects on strategy variability in each group was conducted as in previous studies reported in this thesis, firstly by calculating correlations of uses of strategies across sessions. Table 27, Table 28, and Table 29 show correlations of the most commonly used strategies across sessions.

Table 27:Correlations of uses of relative height across sessions by achievement group and gender (study 5)

	Low Achievement Group		High Achievement Group	
	Boys	Girls	Boys	Girls
Pearson's correlation coefficient	0.128	0.081	0.310	0.355
Significance	0.691	0.725	0.183	0.046

Table 28: Correlations of uses of relative steepness across sessions for high achievement group by gender (study 5)

	High Achievement Group	
	Boys	Girls
Pearson's correlation coefficient	0.377	0.478
Significance	0.101	0.006

Table 29: Correlations of uses of calculation across sessions for high achievement group by gender (study 5)

	High Achievement Group	
	Boys	Girls
Pearson's correlation coefficient	0.480	0.627
Significance	0.032	0.000

Fisher transformations show that in no case are correlations significantly different for boys and girls. (low achievement, height: $z=0.116$, $p>0.05$; high achievement, height: $z=0.166$, $p>0.05$; high achievement, steepness: $z=0.966$, $p>0.05$; high achievement, calculation: $z=0.699$, $p>0.05$).

The second method used to determine differences in variability was by finding for each participant the proportion of strategies in the second session matching a strategy used in the first session. If the proportion of matching strategies was greater than or equal to 0.75, then a participant was considered to have shown low variability. If a participant's proportion of matching strategies was less than 0.75, then that participant was considered to have shown high variability. Chi-squared tests showed no association between gender and level of strategy variability for either the high achievement group ($\chi^2=0.008$, $p>0.05$) or the low achievement group ($\chi^2=0.041$, $p>0.05$).

8.3.3 Patterns of strategy use

Figure 11 and Figure 12 show the distribution of boys and girls across patterns of strategy use in the low and high achievement groups. The expected number of participants in each cell was too small for a chi-square test to be conducted, but some interesting differences between the achievement groups can be seen. These will be discussed in the next section.

Figure 11: Graph to show distribution of boys and girls in high achievement group across patterns of strategy use (study 5)

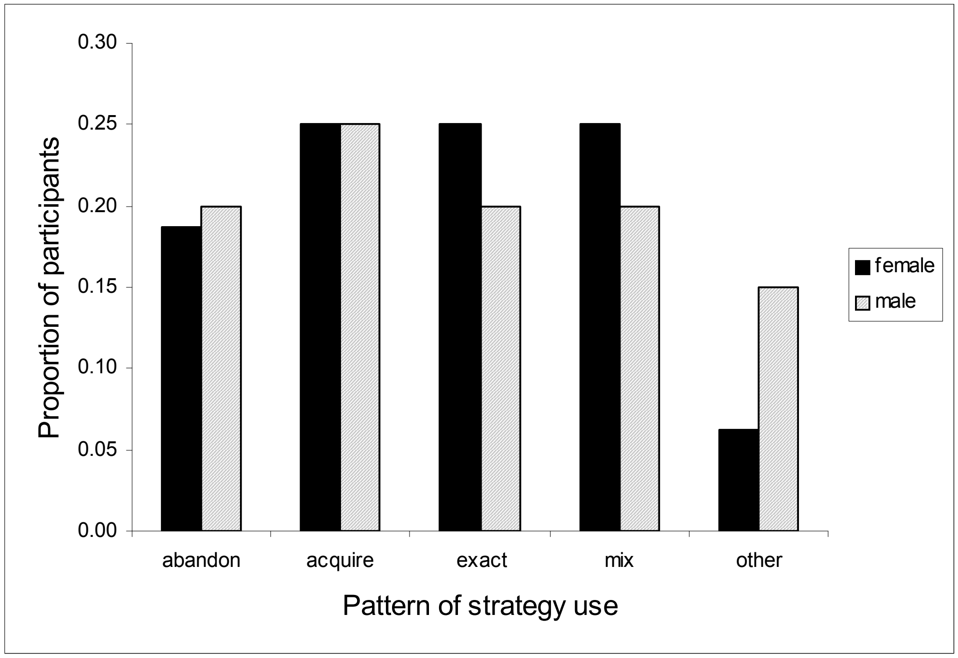
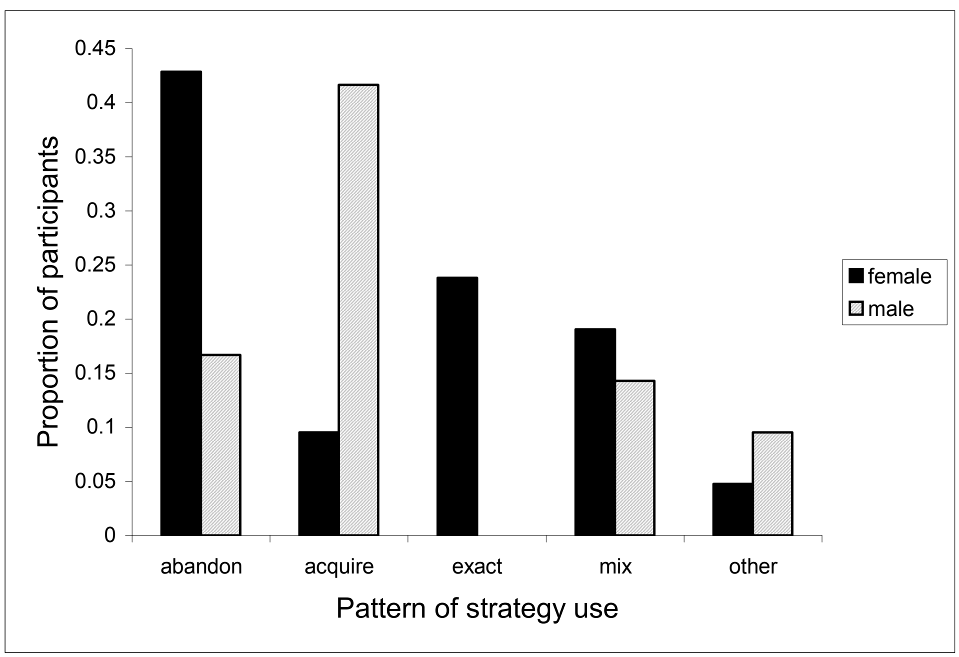


Figure 12: Graph to show distribution of boys and girls in low achievement group across patterns of strategy use (study 5)



8.3.4 Summary

The results show that, as predicted, there is little or no association between variability and gender in either the high- or low-achievement groups.

There are interesting differences in strategies used and in patterns of strategy use over the two sessions to be discussed. In particular, a transition from use of relative heights to use of calculation or comparison of the steepness of lines seems to be represented quite clearly here.

8.4 Discussion

The key aim of this study was to investigate the nature of the results so far discussed in this thesis with respect to the achievement level of the participants involved. There has been an implicit assumption in this thesis so far that the effects observed occur as a result of children's understanding of the problem sets under investigation being sufficient for access to the problems but insufficient to ensure confidence in a correct answer. This study was designed to show that the gender effect on strategy variability is restricted to this period of transition from a state of no understanding to a state of complete understanding by demonstrating children's behaviour on either side of that transition.

The results show the distribution of strategies to be a very good way of determining children's location in the transition of understanding of this problem set. The distribution of strategies used is entirely different for children at the start, middle and end of the transition from lack of understanding to facilitation with the problem set under investigation. The changes in distribution are as expected, as will be discussed below.

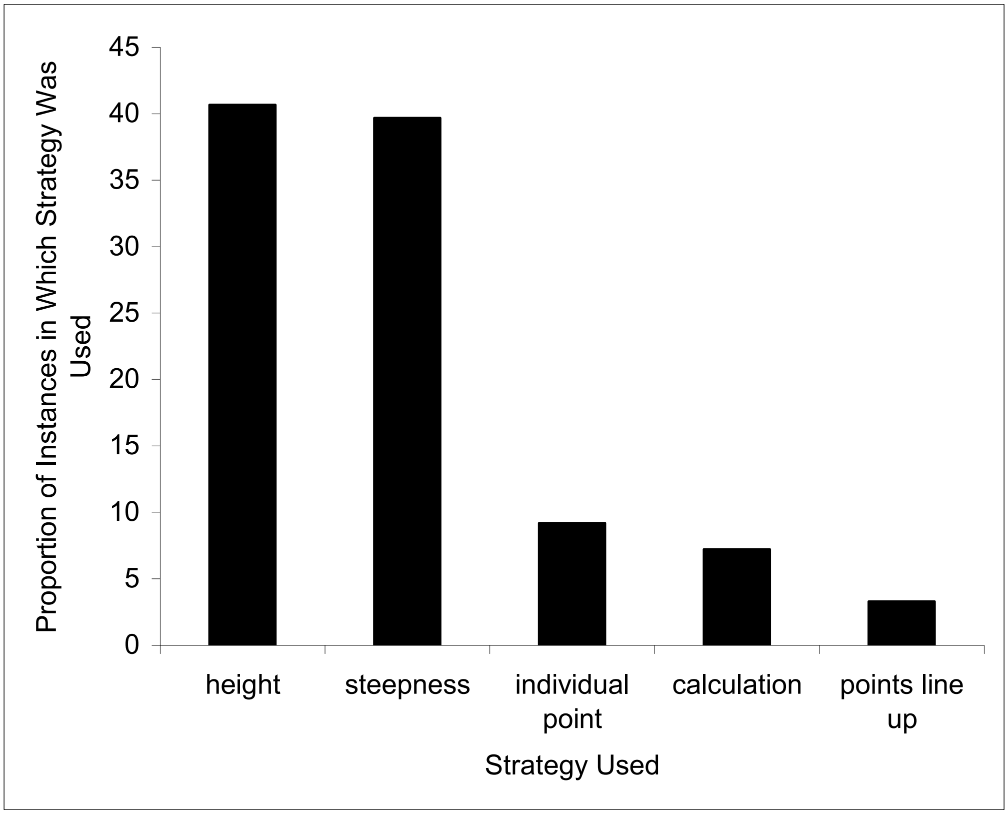
Children in the lower achievement group predominantly used strategies involving the relative height of the two lines. Of all instances of strategy use observed, use of relative height accounted for 53%, with no other strategy being used more than 12% of the time. This backs up the assertion made in the introduction to this study that children at this level in mathematics are not able to engage with this problem set.

Children in the higher achievement group predominantly used strategies involving the relative gradient of the two lines either by visual comparison (44% of strategies used) or the calculation of the step-size between points on the x-axis (26% of strategies used). Significantly fewer children in this group are using the relative heights of lines to inform their judgements of rate of change than in the lower achievement group.

There is a clear shift in distribution of strategy use, from the use of relative heights to the use of gradient, between children in Set 3 and children in Set 1. For comparison, Figure 13 shows the distribution of strategies observed in study 3. This study has shown that the groups of children involved in the studies reported previously in this thesis are at an important stage in their development of conceptual understanding of the rate of change problem set. In each of the previously reported studies in this thesis the proportion of strategies involving relative heights and relative gradients were roughly equal. This indicates that participants were at a point at which they were undergoing a transition in conceptual understanding of the problems.

It has been established then that the children involved in previous studies were at an important stage in conceptual development regarding the rate of change problem set. Is the gender effect on strategy variability only evident at the point of transition? The results of the present study show that there is no effect of gender on strategy variability for either the higher achievement group or the lower achievement group. The reason for this is likely to be the lack of breadth in the distribution of strategies used. It seems that there is a substantially smaller level of variation in strategy variability either side of the point of transition, than there is at the point of transition.

Figure 13: Graph to show distribution of strategies used by participants in study 3



This is an important finding, as it places the gender differences in strategy variability at the point of transition in understanding of a concept. This means that the difference in strategy variability is only apparent at the point of learning and therefore extremely relevant for future consideration of individual differences in mathematical development.

One interesting issue that demands consideration here concerns the differences between the children in each of the groups studied. So far in this section, it has been assumed that the sampling of children of varying achievement levels in mathematics is the equivalent of sampling a group of children at a single achievement level at varying points in time. Unfortunately, this is a problematic assumption to make. None of the children in any of the groups discussed above were formally taught about making judgements about rate of change on the basis of linear graphs. However, the groups show very different approaches to solving the problems. So what is it about these three groups that causes them to behave in different ways?

There are two possible explanations for the three groups' differing behaviour. One of these are compatible with the assumption described above, that sampling differing levels of achievement is equivalent to sampling one level of achievement at different points in time. This explanation is that the three groups described here differ in their speed of mathematical development. The children in set 1 are in set 1 because they are predisposed to learn new mathematical concepts more quickly than children in set 2, who in turn learn new concepts more quickly than children in set 3. If this is the case, then the assumption made above is valid. The alternative explanation, however, is incompatible with that assumption and consists in children differing in achievement in mathematics because of learning new concepts in mathematics in qualitatively different ways. It may be, for example, that students in lower achievement groups in schools achieve at lower levels because of an inappropriate fixation on a particular strategy for a set of problems, as in this case where children were drawn to using relative heights of lines because the problem set looked sufficiently like one in which the use of relative heights could be appropriate. Children in higher achievement groups may be in those groups because of an ability to try out new strategies for problems, or because of an ability to recognise some small differences between problems than indicate the need for new strategies to be used.

It seems that, where the key findings under discussion are related to differences in the ways that children learn new concepts in mathematics, it cannot be assumed that such differences are independent of achievement level. While the results presented here suggest fairly strongly that the gender effect is localised at the point of transition in understanding, the possibility that there is something about children at the particular level of achievement investigated so far that means they show differences in strategy variability, cannot be ruled out completely.

There are two ways in which this problem could be solved. One possibility would be to conduct a longitudinal study aiming to show fluctuations in gender differences in strategy variability during the course of children's transition from a lack of understanding to facilitation with a concept. A second possibility would be to show that problem sets can be generated for children at varying levels of achievement such that these groups show similar gender effects to those discussed in this thesis. These ideas will be discussed further in section 9.3.

Despite the issue raised above, with these findings, the case for the importance of the gender effect is strengthened, for both practical and theoretical reasons. In theoretical terms, previous assertions that high levels of strategy variability are associated with the incidence of conceptual change may need to be refined to account for significant differences between children. In practical terms, different considerations regarding appropriate support and intervention in children’s learning may be warranted at times of transition.

9 General Discussion

9.1 Gender Differences in Strategy Variability

The aim of this project reported in this thesis was to identify one or more group factors that could account for some of the variance in children's levels of strategy variability. The reason for this aim was to help explain why children seem to develop their understanding of mathematical concepts in quite different ways, despite being exposed to the same lesson materials and problem stimuli. In the literature on mathematical development, strategy variability has been shown to be an important factor in children's processes of mathematical development and recent research has identified the existence of individual differences in strategy variability. If the variance in levels of strategy variability amongst children can be better understood, we may be able to come closer to effectively targeting appropriate support and intervention in the classroom to best aid children's developing understanding of mathematics.

The aim of the investigation has been achieved, in that gender has been found to be a strong predictor of strategy variability. The main findings reported in this thesis concern that association between gender and strategy variability. In this section, a description of that association will be derived, on the basis of a combination of the literature discussed in section 4.3 and the evidence obtained during each of the studies reported in this thesis. There will also be some discussion of the potential implications of the findings presented here.

9.1.1 Background

The first two studies reported in this thesis were designed with a number of potential candidates for predictors of strategy variability in mind. The accumulation of candidates was conducted through consideration of those factors that had been reported in the literature to have an association with differences in strategy use. Of the candidates considered, gender was the factor with the greatest weight of evidence behind it. A number of studies have shown that boys and girls favour different solution strategies in response to various problem situations. While most of these involved samples of young children (Carr & Davis, 2001; Carr & Jessup, 1997; Carr et al., 1999; Davis & Carr, 2001; Fennema et al., 1998), some have found differences in strategy use amongst older children, (A. M. Gallagher & de Lisi, 1994; A. M. Gallagher et al., 2000).

The literature suggests that boys are more likely to show adaptivity in strategy use (A. M. Gallagher & de Lisi, 1994; A. M. Gallagher et al., 2000) in that they are more likely to successfully match a suitable strategy to a given problem. Also, boys are less likely to use conventional, rote-learned strategies than are girls (Fennema et al., 1998). Sources of these differences in strategy use have been shown to include children's perceptions of the social impact of strategy choice (Carr & Jessup, 1997), their temperament (Davis & Carr, 2001) and the influence of parents and teachers (Carr et al., 1999).

9.1.2 The effect of gender on strategy variability

The first study reported in this thesis showed that there was a gender difference in strategy variability. The study involved children solving two sets of problems, in two sessions one week apart. The problems asked children to make judgements about rates of change on the basis of straight-line graphs. These were problems that children had access to, in that they understood both the question and the kind of answer they were expected to give, but for which they did not have formal classroom experience. Therefore, in the first session, children were coming up with new strategies in order to arrive at solutions.

A comparison of the strategy sets used by children to solve the problems in each of the two sessions showed that girls and boys behaved differently, showing different patterns of strategy use. Boys were more likely to use a different set of strategies in the second session to that used in the first, while girls were more likely to use a similar set of strategies to the set used in the first session. The association between gender and strategy variability was shown to be independent of strategy distribution, with boys and girls using similar strategies, in similar proportions, to solve the problems in each session. The gender effect was also independent of correctness – whether children's set of strategies in the first session led to correct or incorrect answers, boys were more likely to change the strategies they used and girls more likely to use the same set.

The first study was limited in terms of the depth of data generated for each participant. In order to sample a large number of students in a short amount of time, children were asked to solve the problem sets as whole classes, in examination conditions. As a result of this procedure, there was some loss of data, as a number of children either gave incomplete explanations for their answers or left them out altogether. The second study was designed in order to provide further confidence in the findings of the first. This was achieved by asking a new sample of participants to answer the same problem sets in an individual setting. The one-on-one setting allowed for the experimenter to ensure that participants gave an answer for each question, and gave an explanation for each answer. This allowed a much more thorough and reliable analysis of the strategies used and the patterns of strategy use shown by students across sessions.

The second study produced very similar findings to the first. Boys showed high strategy variability, compared with girls' low variability. In addition, it was possible to compare girls' and boys' strategy use across sessions with a finer degree of detail, by categorising all possible patterns of strategy use and comparing distributions. Boys were most likely to use a completely new set of strategies in the second session, compared to the first. Girls were most likely to reuse at least some of the strategies used in the first session.

With the analysis of the results of the second study, it was established that there was a very interesting association between gender and strategy variability. The next stage of the investigation involved a clarification of the observed gender effect on strategy variability. This was accomplished by considering some potential covariants with gender that might account for the effect.

9.1.3 Ruling out covariants

The third study was designed in order to clarify the observed effect of gender on strategy variability. Gender is known to co vary with a number of cognitive factors that are likely to be involved in problem solving. These include memory, calculation, mental rotation and anxiety, as discussed in section 2.5.3. Some covariants were considered to have a potential role in the association between gender and strategy variability, most importantly, children's memory for either the problem set of strategy set from the first session, their attitude and affect in relation to the problems and their perceptions of contradiction between strategies.

The design and procedure of study 3 was very similar to that of study 2. The only difference was the addition of think-aloud protocols and structured interviews, intended to generate detailed data on the potential covariants described above.

Children's memory for the problem set and strategy set from the first session was assessed during an interview conducted at the end of the second session, as were their perception of contradictions between their two sets of strategies. Memory for the problem set encountered in the first session was assessed by questioning children about similarities and differences between the two sets of problems – the differences were in the context of the problems (either abstract or realistic). Children's attitude to the problems, their temperament and anxiety, were assessed by measuring the amount of time children spent solving each problem. While not a measure with a fine level of detail, analysis of time on task was considered to be a way to detect a range of potential causes of the effect on strategy variability. While time on task may not have been sufficient to pinpoint a particular cause, should some association have been found between time on task and either gender or level of strategy variability, that would have provided motivation for further investigation at a finer level of detail.

The results of the third study were interesting not only for the light shed on the issues just described but for the findings related to the introduction of the think-aloud protocols. The gender effect on strategy variability was found in the third study to be the reverse of the effect observed in each of the two previous studies. In study 3, girls showed higher variability and boys showed low variability. This reactivity to think-alouds will be discussed at greater length in the next section.

With reference to the potential covariants with gender that study 3 was intended to investigate, there was no evidence that any of the covariants considered could be considered mediators of the association between gender and strategy variability. Children's memory for the problem set from the first session was related to neither gender nor strategy variability. This was surprising for two reasons, as memory has often been shown to be related to gender (e.g. Herlitz et al., 1997; Maccoby & Jacklin, 1974) and any differences in memory for the first session were expected to have a considerable effect on children's use of strategy in the second session. There was sufficient variability in children's memory for the first session that any association with either gender or strategy variability would have been detected. The effect of children's memory for the strategy set used in the first session was not possible to analyse as children uniformly reported the strategy set used in the first session to be identical with that used in the second session.

Children's perceptions of contradiction were fairly accurate. Where children were confronted with a contradiction in answers given in their two completed problem sets, the vast majority were quick to respond by changing one or other answer so that they matched and so no longer contradicted one another. Those children who did not want to 'fix' their contradictory answers argued their case by referring to the wording of the question rather than the mathematics of the problem. The five children concerned were evenly distributed across gender.

Time on task was analysed with respect to its potential association with both gender and strategy variability. The amount of time was thought to have some potential association with gender, due to the results of Davis and Carr (2001) which showed that boys' strategy choices were affected by their level of impulsivity, while girls' were affected by their level of inhibition. Also it was considered that time on task should be a good indicator of children's levels of anxiety, a further factor shown in the literature to be related to gender, with the potential to affect strategy variability through its connection with working memory (Ashcraft & Kirk, 2001; Miller & Bichsel, 2004).

There was no significant association between time on task and either gender or strategy variability. This again was surprising, as the analysis of time on task was intended to be a way of testing for a number of possible covariants with gender that might be affecting strategy variability. As no association was found between time on task and either gender or variability, it seems that a number of potentially mediating factors can be ruled out.

It therefore remains to be seen whether some factor or factors can be found that can explain the relationship observed in each of the studies reported in this thesis between gender and strategy variability. Some possible options for future exploration in this area are discussed in the next section.

9.1.4 Establishing a locus

The last two studies reported in this thesis both have the aim of describing the locus of the association between gender and strategy variability. It is important that it is possible to say as precisely as possible which children solving which problems show this association, and to what extent any particular level of strategy variability is a durable characteristic of a child.

Study 4 was designed in order to investigate the durability of strategy variability in children. This was achieved by repeating the procedure of study 3 with the same children who participated in study 3, six months later. The children still had not encountered the problems involved in the problem sets in the classroom. An additional benefit of the study would be an opportunity to re-examine the gender effect on strategy variability, especially important due to the reversal of the effect in study 3 compared to that observed in the first two studies.

Study 4 showed the gender effect on variability to have a similar strength to that found in study 3, in the same direction. The findings of study 4 also indicated that children's level of strategy variability had a high level of consistency across the six month period between tests.

These were valuable findings, as they suggest that further research in this area is likely to be profitable in terms of practical effects in the classroom. Some issues connected with this will be discussed below and in the next section.

The aim of the fifth study reported in this thesis was to investigate the extent to which the relationship between the ability of children and the difficulty of the problem set determined the size of the gender effect on variability. This study reused the design and procedure of study 1 in order to accumulate a large quantity of data, as the detailed data produced using think-aloud protocols and interviews was unnecessary. The children participating in study 5 had levels of achievement in mathematics both above and below that of the groups tested in studies one to four. The low achievement group were expected by their teacher to achieve a level 4 or below in the mathematics SAT, while the high achievement group were expected to achieve a level 6 or above.

It was expected that the gender effect on strategy variability was particular to a certain stage of development in children's conceptual understanding of the problem set under investigation and therefore would not be seen in children either side of this point. Throughout each of the earlier studies reported in this thesis, the participating children were at a point at which they had access to a wide range of strategies for generating answers to the given problems, but did not have a thorough understanding of the mathematical concepts involved or a very sophisticated method of deciding on a particular strategy. The aim of study 5 was to show that the gender effect was not observable outside of this group of children and to show how children enter and leave this transition point.

As expected, the results showed no evidence of a gender effect on strategy variability in either the high- or the low-achievement group. The comparison between the two groups in terms of the distribution of strategies shown by children proved to be very interesting. The low-achievement group predominantly used the relative height of the lines to answer the given problems. This is a naïve strategy whose use is due to children's experience of using the height of lines to answer question regarding absolute values rather than derivatives. Over 50% of questions seen by the participants in the low achievement group were answered using a strategy involving the relative heights of the lines and no other strategy was used on more than 12% of questions. The fact that so many questions were answered using this one strategy explains why the low achievement group showed no association between gender and strategy variability – there was not sufficient variance in strategy use for any factor to show an association.

The high achievement group predominantly used appropriate strategies for generating solution to the questions set. The most common strategy used by this group involved the relative gradients of the two lines – over 40% of questions seen by this group were answered using this strategy. The second most common strategy used by this group, used to answer over 25% of questions, involved some calculation that provided an equivalent to the gradients of the two lines. Again, similarly to the low achievement group, the reason that no association was found between gender and strategy variability was due to the lack of variance in strategies used.

The transition from a lack of understanding to a more thorough understanding of the problems was clear from a comparison of the distributions of strategies used by the two groups in study 5. Although these data were derived from a cross-section of achievement groups rather than a longitudinal study of a group of children, it has been assumed that the development of the relevant concepts is equivalent in the two situations. The data suggest that the transition in conceptual understanding of these rate of change problems involves three stages. The first stage involves children learning that strategies involving the relative heights of lines are not appropriate for solving this problem. The second stage involves children experimenting with a wide range of strategies that generate a solution to the problems. This second stage involves high levels of variability in strategy use. The third stage involves children settling into a stable pattern of using one particular strategy that can be relied upon to generate correct answers.

The gender effect on strategy variability can only be observed during the second stage of the transition in understanding, as this is the stage during which children are more likely to be experimenting with new strategies for solving problems. The data from study 5 then correspond very well with the assertions of Siegler (1995), which suggests that conceptual change is associated with an expansion followed by a contraction of the strategy repertoire.

9.1.5 Implications of these findings

There are two sets of implications to be considered here. These are practical and theoretical. Firstly, the findings described above will be considered in terms of their implications for theories of mathematical development. Secondly, the potential impact on the mathematics classroom will be discussed.

The coincidence of strategy variability with conceptual change is a key feature of some key theories of development, including connectionist, information-processing and dynamic systems accounts (see section 2.2). It is clear then that any differences between children in terms of level of strategy variability at the time of conceptual change are ripe for investigation and will play a large part in the development of future theories of development.

It is only fairly recently that individual differences in strategy variability have begun to be researched (e.g. Schunn & Reder, 1998). The major addition that the work reported in this thesis has made to the field has been the demonstration that a significant portion of individual differences in strategy variability can be accounted for by gender. It is becoming increasingly clear that children cannot be considered to form homologous groups of learners whose development follows a simple trajectory. Research in mathematical development was problematised in Siegler (1987) where findings indicated that averaging data across trials gave a misleading picture of strategy use. It now seems that research in mathematical development can further be problematised with the assertion that averaging data over participants will give a misleading picture of development.

In each of the first three studies reported here, analysis of results from the sample as a whole would have produced findings inconsistent with data from either the male or female population alone. Indeed, the findings of Mevarech and Stern (1997) were replicated in study 1 only for the male portion of the sample. Analysis of the data from female participants and of sample as a whole showed no significant association between the order of contexts presented and improvement between sessions, whereas analysis of the data from male participants only showed a fairly strong association. It seems increasingly likely that if individual and group differences are neglected in research in children's learning, then an incomplete and possibly misleading picture will result.

The findings discussed above also demonstrate quite strongly the fact that development is a complex and dynamic process, during which various factors can have more or less effect on events over time. This may be an issue for information-processing theories and attempts to model student behaviour mathematically. Mathematical models are very good at describing situations with relatively few degrees of freedom, but the modelling process may prove to be problematic where there are group and individual factors at work that alter and disrupt the learning process to any great extent.

Differentiation is an important part of classroom mathematics teaching. The findings presented here will, with further work, help to provide a valuable means of providing the most appropriate forms of support and intervention to children in their learning of mathematics. This investigation has shown that children can differ substantially in the ways that they develop new strategies to solve problems, and therefore develop new conceptions of mathematical objects.

The further work required before these findings become usable in the classroom will be discussed below in section 9.3. However it is possible to see already how teaching and learning could benefit. For example, research like that presented here is making more and more clear the picture of how children come to develop new conceptions in mathematics, and this can help practitioners to determine when are the right times to present new ideas to students. One thing that has been relatively constant in theories of learning over the last few decades has been the idea that children should be presented with new challenges at a time when they are within their conceptual grasp, but outside their procedural facility. As early as 1912, Thorndike imagined a textbook for children in which pages would only become visible to the reader once earlier pages had been read and understood (Thorndike, 1912, p.165). Wood, Bruner and Ross (1976) describes a study in which children's learning was greatest when their mothers most closely matched their support to their children's abilities. Of course there is also Vygotsky's Zone of Proximal Development (Vygotsky, 1978), already mentioned above, which is the set of problems that a child cannot yet solve independently, but could solve with the aid of a peer or adult. Vygotsky claimed that teaching was most effective when within a given child's ZPD.

The findings presented here can be utilised with the above ideas in mind. The results of study 5, for example, help to pinpoint a good time for the introduction of rate of change problems in the classroom. Children in the low achievement group were probably not ready for this material; they showed low strategy variability and predominantly used a naïve strategy based on a strategy for a simpler class of problem. The high achievement group were probably past the point at which introduction of rate of change as a concept would be most useful. This group again showed low variability, and predominantly but not exclusively used appropriate strategies for the problems. The intermediate achievement group were probably at the most suitable point for the introduction of material on rates of change. They showed the highest level of variability and the widest range of strategies used. As many students used inappropriate strategies as appropriate strategies. If material can be presented at a time a close as possible to the point at which children are ready to learn about it, then efficiency of learning should be maximised. This investigation goes some way at least to support this goal.

In addition to the timing of introductions to new material, a second way in which these findings could be used in the classroom is in determining the kinds of support and intervention required by children in their individual mathematical development. The studies reported here have shown that children do not all develop new strategies in response to problems in the same way. It has also been demonstrated that strategy variability is a fairly durable characteristic of children's strategy development behaviour. Further work may help to establish differential teaching strategies for children showing high and low strategy variability in response to problems.

9.2 Reactivity to think-alouds

A surprising secondary major finding presented in this thesis is the reactivity to think-aloud protocols shown by comparison of studies 2 and 3. The only procedural difference between the two studies was the addition of think-aloud protocols in study 3. In both studies, participants were at the same level of achievement (level 5 on Year 9 mathematics SAT) had the same experience of the problem set used and were asked to complete the problem set individually. Identical problem sets were used in each study.

Different participants were involved in studies 2 and 3, however. Also the four mathematics teachers from whose classes participants were sampled for study 3, were different to the four teachers from whose classes participants were sampled for study 2. It is possible that there were sufficient differences in either teaching methods or participant behaviour to cause the crossover effect observed between studies 2 and 3. However, this seems unlikely, given the analysis of results across the two studies. Firstly, while participants' classroom teachers differed between studies 2 and 3, 2 of out 4 classes that participated in study 3 were from the same school as participated in study 2. Secondly, there were no significant differences in strategy variability between schools or between classes for either boys or girls in either study 2 or study 3. While this is not conclusive proof that verbal reports are the only explanation for the observed differences in participant behaviour between study 2 and 3, the probability that the differences were due to either biased samples, or different teaching methods is very low.

The difference in the results in the two studies was unexpected. There is a fairly substantial literature that suggests there should be no reactivity to think-aloud protocols (e.g. Ericsson & Simon, 1993; Kirk & Ashcraft, 2001; Kuusela & Paul, 2000; Robinson, 2001; Wilson, 1994). The gender effect observed in studies 1 and 2 was completely reversed in study 3. Not only was there a significant difference in strategy variability between boys and girls in both study 2 and study 3, but there was a significant difference in strategy variability between boys in study 2 and boys in study 3 and between girls in study 2 and girls in study 3. It seems that the only explanation for these findings is that children's problem solving behaviour in the second session changed due to the think-aloud protocols used during the first session.

It is important to note here that the reactivity was only apparent in children's patterns of strategy use across session and in their levels of strategy variability, compared with those found in studies without think-alouds. Other factors, such as the distribution of strategies used and the number of correct answers were unaffected. Without the comparison of analyses of strategies across sessions, with and without think-alouds, the reactivity would have been impossible to detect.

There are two main implications of the observed reactivity to think-aloud protocols. The first is that it shows that the gender effect on strategy variability, while fairly robust, appears to be somewhat dependent on the context in which problem solving takes place and/or concurrent demands on children's cognitive resources. The second is that without the use of a control group in which think-alouds are not used, the use of think-alouds in investigation of strategy use appears to involve a substantial risk of producing incomplete or misleading results. For example, if think-alouds had been used in this investigation from study 1, the conclusions drawn would have been quite different to those drawn here.

9.3 Directions for Future Research

The main finding of this thesis has been that research in the area of individual differences in strategy variability is likely to be interesting and worthwhile. This section then takes on a great deal of importance in suggesting ways in which the findings discussed earlier might be developed in the future. The aim of this section is to demonstrate how this thesis can be thought of as a large part of the foundation for a future body of work that will help to explain a great deal of the variation that can be observed in children's achievements in mathematics classrooms.

There are a few important questions that will need to be addressed initially. These are questions regarding the generalisability of the findings presented here. The studies reported in this thesis all share the same problem set, involving rate of change problems with linear graphs. On the basis of the literature review of this thesis is seems that similar results to those presented here will be found in future studies involving other problem sets. Initially, it will probably be important to limit changes to the problem set in order to be able to accurately assess any changes in behaviour due to the problems presented. However, given the assessment of the research on graph problems in the literature review of this thesis, it is likely that similar findings will be obtained in studies involving a wide variety of potential problem sets in the mathematical domain, where problems are within children's ZPD and where there is a breadth of potential solution strategies. Examples of potential problem domain might include the solution of algebraic equations, of geometry problems, of arithmetic problems and possibly various others. Data that replicates the findings reported in this thesis, but with the use of different problem sets will be an extremely valuable step towards a coherent body of research on individual differences in strategy variability.

In addition to the problem set, a second constant has been the age and ability of the participants. In the last study reported here, the ability level of the children was varied, but this was without a corresponding adjustment of the difficulty of the problem set, and therefore confirmed only that the effects described earlier in the thesis applied to children's solution of problems within the ZPD. It seems likely that if the problem set was varied with participants' level of ability so that it fell within the ZPD, then similar effects to those reported in this thesis would be observed. Again, data supporting this assertion will be valuable in building a foundation for a coherent body of work.

With regard to the final study reported in this thesis, concerning the localisation of the gender effect at the point of transition in understanding, there seems to be further work required to clarify some of the issues. The main problem with the results as they stand is the assumption that sampling children at varying ability levels is equivalent to sampling a group of children at various stages in their development. To solve this problem, a longitudinal study could be designed that followed children from a point at which they have limited understanding of a concept to a point where they have some facilitation with that same concept. Strategy variability within and between sessions would be analysed to determine how this fluctuates about the point of learning of the new concept. The difficulty with this would be the scale of data collection required. Firstly, the sample size needed for making judgements about gender differences in strategy use is relatively large. Secondly, it would be difficult to know in advance how long the transition from lack of understanding to facilitation with a new concept will take.

An alternative study would aim to solve the problem in a more efficient way. This would involve the generation of new problem sets such that groups at differing levels of achievement could be matched with problem sets to which children have access without facilitation. If it were possible to show the gender effect on strategy variability for children at different levels of achievement, using different problem sets, then it would be reasonable to conclude that gender differences are present at the point of learning a new concept in mathematics.

This thesis has achieved its initial aim, of determining a group factor that accounts for differences in strategy variability, in showing that gender is a strong predictor of variability over time. It was made clear in the introduction that there are likely to be a number of group factors that contribute to individual differences in strategy variability. It will be interesting now to investigate some of the other potentially predictive group factors. Two factors have been mentioned already in this section, although with a slightly different slant. The age and ability of children can both be considered as group factors that might help to explain individual differences in strategy variability. Studies conducted in order to test the generalisability of the gender effect reported here across age and ability will therefore have the benefit of being able to show either that the gender affect generalises over age and ability or that age and ability account for some of the variation across individuals themselves.

Alternative group factors suitable for further investigation might include race and socio-economic status (as discussed in section 2.4.2). These factors, along with gender, have an evidence base for predictive power for level of achievement in mathematics. However, there is little, if any, evidence for an association with differences in strategy use.

All findings regarding individual and group differences in strategy variability are expected, in the long term, to provide information that will enhance the effectiveness of classroom mathematics teaching. It is likely that children showing high and low variability will have different needs for support and intervention in their learning of new mathematical concepts. The end goal of research in this area must be to develop practical methods for teaching mathematics to children in ways that meet their individual needs. Meeting the needs of children on the basis of their individual level of strategy variability is probably a number of years away but is nevertheless a worthwhile goal.

A second finding of the thesis has concerned participants' reactivity to think-alouds, observed through comparison of studies reported here that have differed only in term of the inclusion or not of think-aloud protocols and retrospective reports. Of course, as discussed previously, it can not be determined whether children's reactivity was due to the think-alouds, the retrospective reports or a combination of the two, but as the use of think-alouds with retrospective reports is fairly standard practice, the finding is an important one. It is essential that researchers making use of think-aloud protocols have confidence in the reliability of their methods. If the use of think-alouds can be shown to affect children's problem-solving behaviour in future trials, then that raises large issues of reliability.

A recommendation on the basis of the studies presented in this thesis is that future research using think-aloud protocols for investigating strategy use over time should make use of a control group in which think-alouds are not used. There is enough data available without analysis of think-alouds such that a decision can be made as to the effect of think-alouds on patterns of behaviour across trials. If children exhibit different patterns of strategy use under think-aloud condition than they do without, then the data produced through the think-alouds is likely not telling us about the situation we think it is. Studies of this kind will enable both more reliable research regarding patterns of strategy use and also the generation of more information as to the appropriate uses of think-alouds and retrospective reports for investigating patterns of strategy use over time.

10 Conclusion

In studies 1, 2 and 3, it was shown that there was a clear difference in strategy variability between boys and girls. In study 4, it was shown that strategy variability is a stable characteristic; children's levels of strategy variability measured six months apart showed a significant positive correlation. Study 5 showed that these findings are applicable to problems located in the zone of proximal development.

10.1 Gender Difference in Strategy Variability

The aim presented at the beginning of this thesis was to identify at least one group factor that influenced strategy variability. This has been achieved. The first major conclusion of this thesis is that there is a gender difference in patterns of strategy use in response to rate-of-change problems. The direction of this difference depends on other factors, as will be described below, but the existence of the difference is constant throughout all the studies conducted as part of this thesis.

This finding is important because it shows that there is a stable and durable characteristic of children that plays a part in determining strategy variability in response to rate-of-change problems. This characteristic is certainly associated with gender and potentially associated with other factors, yet to be investigated. The effect of this will be to open up a number of research questions regarding individual children's strategy variability, its causes and its effects. The importance of individual differences in strategy variability is highlighted by strong indications in the literature that strategy variability is linked with ability to learn new concepts. It will be important for both research and practice in education that answers to these questions of individual difference are pursued.

In addition to providing something of a foundation on which to build future work, the findings presented here are in themselves a significant addition to the gender difference literature. The vast majority of the literature on gender differences in mathematics education addresses static situations. In fact it is only relatively recently that researchers have progressed from simply analysing performance differences on national tests. Even more recent research, which addresses questions regarding specific differences in strategy use, attitude or other cognitive factors, has involved the study of static situations, therefore dealing with performance rather than learning. The position presented in this thesis is that questions of performance are not as interesting or useful as questions of learning, especially in the arena of gender difference. The justification of this position is that gender difference research on *performance* is complicated by the fact that questions of cause are largely insoluble and that implications for the classroom are limited. Findings of differences in *learning* of mathematics, however, although sharing the problem of the identification of causes, at least have directly applicable implications for the classroom. Differences in learning can be associated with differentiable needs for support and intervention in a way that differences in performance cannot.

10.2 Reactivity to think-aloud protocols

The findings of studies 2 and 3 are strong evidence for reactivity to think-aloud protocols. Reactivity to protocols is an extremely important but often overlooked factor on which the validity of a great deal of research depends. Consider the potential conclusion of either study 2 or 3 if conducted independently. Presumably, given the results of study 2, one would conclude that boys show high strategy variability compared to girls' low variability in response to rate-of-change problems presented one week apart. Given the results of study 3, one could conclude with equal confidence that boys show low variability compared with girls' high variability. The problem is clear from this example. The findings of either study, in absence of the other, imply a conclusion that is at best incomplete and at worst misleading.

A large number of papers are published each year that report studies involving the use of think-aloud protocols. There is often an implicit assumption that the use of think-alouds has no effect on participants' cognitive processes or behavioural outcomes. On the basis of the findings presented here, it can be argued that without the use of a control group in which think-aloud protocols are not used, it is impossible in many cases to say for certain that they have no effect on measured behaviour.

This effect would not have been discovered if not for the analysis of *patterns* of problem-solving strategies over time. The comparison of studies 2 and 3 shows no difference in the type or distribution of strategies used by participants either in trial 1 or trial 2. It is only through analysis of the *dynamics* of strategy use that reactivity becomes evident. It is then clear that despite appearances, children's use of strategy in at least trial 2 of study 3 have been influenced by the experimenter's use of think-aloud protocols.

Since there is a good deal of research that involves the use of multiple trials with think-alouds, without a control group of participants not asked for think-alouds, there seems to be a real issue of validity for a large amount of research.

10.3 Interaction between gender and think-aloud protocols

The interaction between gender and think-alouds can perhaps be thought of as a corollary of the above conclusions. If there are gender differences in strategy variability and there is reactivity to think-aloud protocols, then it may not be thought surprising that there is an interaction between the two. However, it is difficult not to be surprised by the nature of the observed interaction.

Studies 2 and 3 can be thought of as one between-groups study with the dependent variable being strategy variability and independent variables gender and think-aloud intervention. This collapsing of study 2 and 3 can be done as a result of the fact that there were no other differences between the two studies, other than the addition of the think-aloud and retrospective report. Thought of in this way, as one between-groups study, the data from studies 2 and 3 give strong evidence for an interesting interaction between gender and think-aloud intervention. There is a striking crossover effect, illustrated in Table 30.

Table 30: Interaction between gender and the use of think-alouds

	Male	**Female**
Think-aloud	Low variability	High variability
No think-aloud	High variability	Low variability

The extent of the effect is beyond any that should be expected given the literature currently available. Even if it were known that gender effects on strategy variability were mediated by use of think-alouds, it would be more natural to assume that students of both genders would move towards the same end of the strategy variability scale as a result. The fact that boys become less variable and girls become more variable with the use of think-alouds suggests that there is a very interesting mechanism at work. Possible causes have been discussed previously and do not need to be repeated here, but it is clear that further research into the nature of this effect will be profitable.

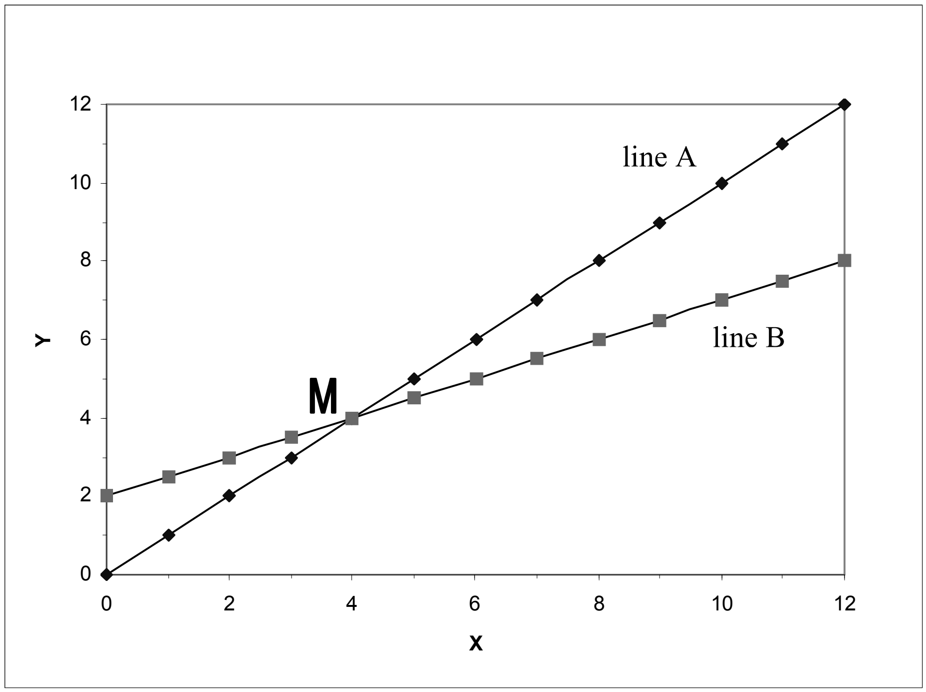
The paradoxical nature of these findings, that the gender difference in strategy variability can be so stable, yet be reversed with the introduction of think-alouds suggests that implications for the classroom are as yet impossible to determine but are vital to pursue.

10.4 Summary

This thesis is a significant addition to the literature on the psychology of mathematics education in three ways. Gender differences have been identified in dynamic situations, validity issues have been identified regarding the use of think-aloud protocols and new research questions have been identified regarding individual differences in strategy variability.

Appendices

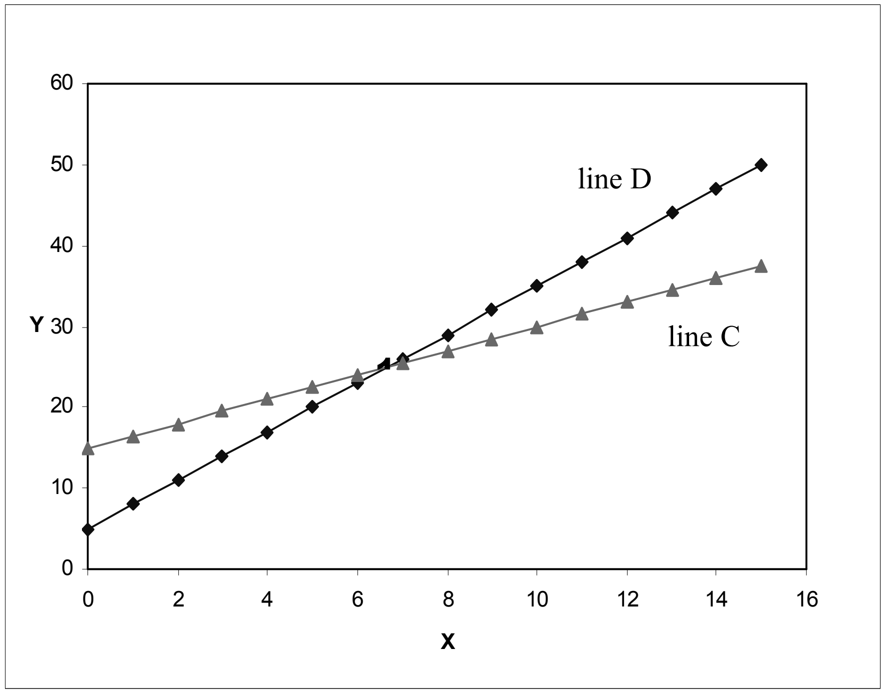
Appendix A - The abstract problem set



1. Up to point M, is line A increasing faster, slower, or at the same rate as line B?

2. At point M, is there a change in the rate of increase of line A?

3. From point M onwards, is line A increasing faster, slower, or at the same rate as line B?



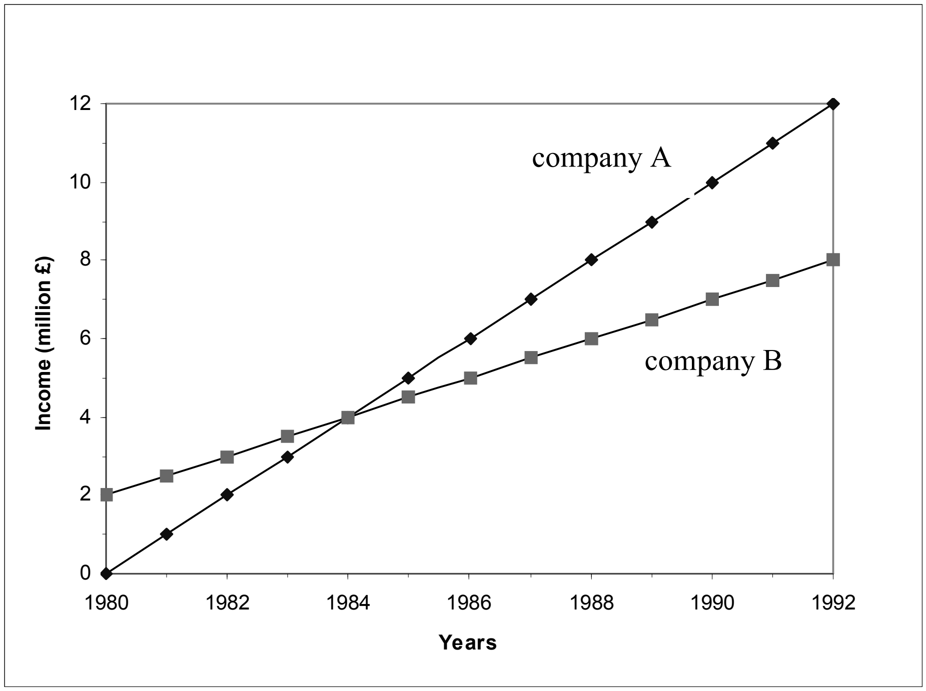
These questions refer to a new graph – shown above:

4. Up to point N, is line C increasing faster, slower, or at the same rate as line D?

5. At point N, is there a change in the rate of increase of line C?

6. From point N onwards, is line C increasing faster, slower, or at the same rate as line D?

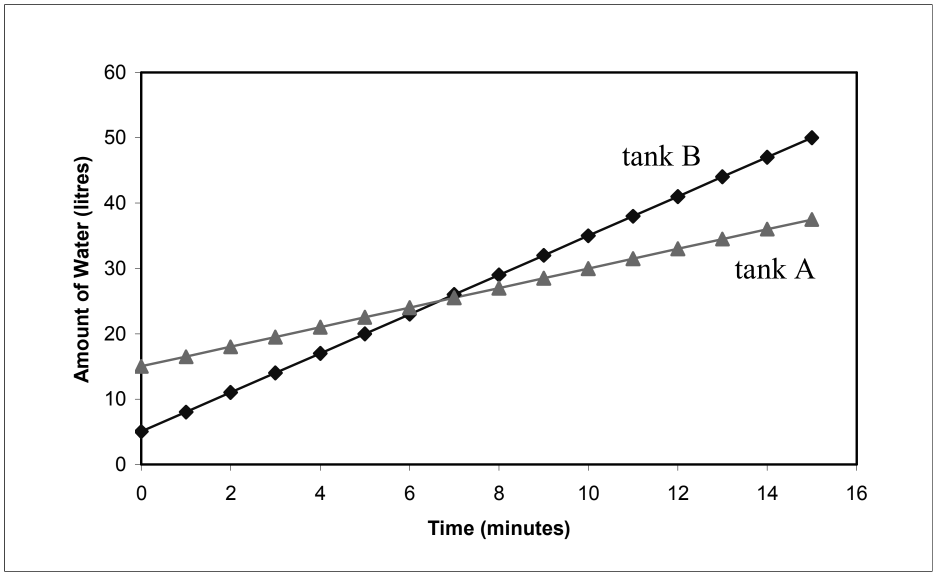
Appendix B - The realistic problem set



1. Before 1984, was the income of company A increasing faster, slower or at the same rate as the income of company B?

2. In the year 1984, was there a change in the rate of increase of income of company A?

3. From the year 1984 on, was the income of company A increasing faster, slower or at the same rate as the income of company B?



This new graph represents the amount of water in two fish tanks (tank A and tank B) as they are being filled up. Use this graph to answer the next three questions:

4. Before 7 minutes, was the amount of water in tank A increasing faster, slower or at the same rate as the amount of water in tank B?

5. At 7 minutes, was there a change in the rate of increase in the amount of water in tank A?

6. From 7 minutes on, was the amount of water in tank A increasing faster, slower or at the same rate as the amount of water in tank B?

References

Alibali, M. W. (1999). How children change their minds: strategy change can be gradual or abrupt. *Developmental Psychology, 35*(1), 127-145.

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25*(4), 468-573.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5-11.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1997). Situative versus cognitive perspectives: Form versus substance. *Educational Researcher, 26*(1), 18-21.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology-General, 130*(2), 224-237.

Branch, J. L. (2000). *The trouble with think alouds: Generating data using concurrent verbal protocols.* Paper presented at the CAIS 2000: Dimensions of a global information science, University of Alberta.

British Educational Research Association. (2004). *Revised Ethical Guidelines for Educational Research*. Macclesfield, Cheshire: BERA.

British Psychological Society. (2000). *Code of Conduct, Ethical Principles and Guidelines*. Leicester: British Psychological Society.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin, 125*(3), 367-383.

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

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology-Applied, 4*(2), 75-100.

Carr, M., & Davis, H. (2001). Gender differences in arithmetic strategy use: A function of skill and preference. *Contemporary Educational Psychology, 26*(3), 330-347.

Carr, M., & Jessup, D. L. (1997). Gender differences in first-grade mathematics strategy use: Social and metacognitive influences. *Journal of Educational Psychology, 89*(2), 318-328.

Carr, M., Jessup, D. L., & Fuller, D. (1999). Gender differences in first-grade mathematics strategy use: Parent and teacher contributions. *Journal for Research in Mathematics Education, 30*(1), 20-46.

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1985). Mathematics in the Streets and in Schools. *British Journal of Developmental Psychology, 3*(MAR), 21-29.

Case, R. (1992). *The Mind's Staircase: Exploring the Conceptual Underpinnings of Children's Thought and Knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (2001). Spatial-mechanical reasoning skills versus mathematics self- confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *Journal for Research in Mathematics Education, 32*(1), 28-57.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23*, 43-71.

Cobb, P., Wood, T., & Yackel, E. (1991). A constructivist approach to second grade mathematics. In E. von Glasersfeld (Ed.), *Radical Constructivism in Mathematics Education*. Dortrecht, The Netherlands: Kluwer Academic Publishers Group.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*(5), 382-393.

Davis, H., & Carr, M. (2001). Gender differences in mathematics strategy use - The influence of temperament. *Learning and Individual Differences, 13*(1), 83-95.

Delgado, A. R., & Prieto, G. (1997). Mental rotation as a mediator for sex-related differences in visualization. *Intelligence, 24*(3), 405-416.

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence, 32*(1), 25-32.

DFES. (2001). *Framework for teaching Mathematics: Years 7, 8 and 9*. London: DFES.

DFES. (2005a). *GCE/VCE A/AS examination results for young people in England, 2003/04*. London: DFES.

DFES. (2005b). *National curriculum assessments at Key Stage 3 and Key Stage 2 to Key Stage 3 value added measures in England 2004/05 (Final)*. London: DFES.

Dowker, A., Flood, A., Griffiths, H., Harriss, L., & Hook, L. (1996). Estimation strategies of four groups. *Mathematical Cognition, 2*(2), 113-135.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis - Verbal Reports as Data*. Cambridge, Mass.: The MIT Press.

Fan, X. T., Chen, M., & Matsumoto, A. R. (1997). Gender differences in mathematics achievement: Findings from the National Education Longitudinal Study of 1988. *Journal of Experimental Education, 65*(3), 229-242.

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.

Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.

Gallagher, A. (1998). Gender and antecedents of performance in mathematics testing. *Teachers College Record, 100*(2), 297-314.

Gallagher, A. M., & de Lisi, R. (1994). Gender differences in scholastic aptitude-test - mathematics problem-solving among high-ability students. *Journal of Educational Psychology, 86*(2), 204-211.

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

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*(18), 1-16.

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences*(19), 229.

Geary, D. C. (1999). Sex differences in mathematical abilities: Commentary on the math-fact retrieval hypothesis. *Contemporary Educational Psychology, 24*(3), 267-274.

Gelman, R., & Gallistel, C. R. (1978). *The Child's Understanding of Number*. London: Harvard University Press.

Gilligan, C. (1982). *In a different voice*. Cambridge, MA: Harvard University Press.

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology, 76*(2), 104-122.

Gorard. (2001). *Quantitative Methods in Educational Research*. London: Continuum.

Greeno, J. G. (1997). Response: On claims that answer the wrong questions. *Educational Researcher, 26*(1), 5-17.

Halpern, D. F. (1986). *Sex differences in cognitive abilities*. Hillsdale, NJ: Erlbaum.

Hart, K. (1981). *Children's Understanding of Mathematics: 11-16*. London: John Murray.

Hedges, L. V., & Nowell, A. (1995). Sex-differences in mental test-scores, variability, and numbers of high-scoring individuals. *Science, 269*(5220), 41-45.

Hembree, R. (1990). The nature, effects and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*(1), 33-46.

Herlitz, A., Nilsson, L. G., & Backman, L. (1997). Gender differences in episodic memory. *Memory & Cognition, 25*(6), 801-811.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance - a metaanalysis. *Psychological Bulletin, 107*(2), 139-155.

Johnson, S. (1996). The contribution of large-scale assessment programmes to research on gender differences. *Educational Research and Evaluation, 2*, 25-49.

Kimura, D. (1999). *Sex and Cognition*. Cambrdge, Mass.: MIT Press.

Kirk, E. P., & Ashcraft, M. H. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology-Learning Memory and Cognition, 27*(1), 157-175.

Kuczaj, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior, 16*(5), 589-600.

Kuusela, H., & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *American Journal of Psychology, 113*(3), 387-404.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

LeFevre, J., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*(1), 216-230.

Lerman, S. (1996). Intersubjectivity in mathematics learning: A challenge to the radical constructivist paradigm? *Journal for Research in Mathematics Education, 27*(2), 133-150.

Lerman, S. (2000). A case of interpretations of social: A response to Steffe and Thompson. *Journal for Research in Mathematics Education, 31*(2), 210-227.

Lussier, G. (1996). Sex and mathematical background as predictors of anxiety and self-efficacy in mathematics. *Psychological Reports, 79*(3), 827-833.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Masters, M. S., & Snaders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics, 23*(4), 337-341.

Mevarech, Z. R., & Stern, E. (1997). Interaction between knowledge and contexts on understanding abstract mathematical concepts. *Journal of Experimental Child Psychology, 65*(1), 68-95.

Miller, H., & Bichsel, J. (2004). Anxiety, working memory, gender, and math performance. *Personality and Individual Differences, 37*(3), 591-606.

Mullis, I. V. S., Martin, M. O., Fierros, E. G., Goldberg, A. L., & Stemler, S. E. (2000). *Gender Differences in Achievement: IEA's Third International*

Mathematics and Science Study. Boston, MA: International Study Center, Lynch School of Education, Boston College.

Nickson, M. (2000). *Teaching and Learning Mathematics*. London: Cassell.

Nisbett, R. E., & DeCamp Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 1977.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math not equal me. *Journal of Personality and Social Psychology, 83*(1), 44-59.

Piaget, J. (1952). *The Child's Conception of Number*. London: Routledge & Kegan Paul Ltd.

Piaget, J. (1977). *The development of thought: equlibration of cognitive structures* (A. Rosin, Trans.). New York: Viking Press.

Piaget, J., & Inhelder, B. (1969). *The Psychology of the Child*. New York: Basic Books.

Raven, J. C. (1976). *Standard Progressive Matrices*. Oxford: Oxford Psychologists Press.

Reis, S. M., & Park, S. (2001). Gender differences in high-achieving students in math and science. *Journal for the Education of the Gifted, 25*(1), 52-73.

Robinson, K. M. (2001). The validity of verbal reports in children's subtraction. *Journal of Educational Psychology, 93*(1), 211-222.

Robson, C. (1993). *Real World Research*. Oxford: Blackwell.

Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. J., & Marchant, H. (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology, 24*(3), 181-266.

Sammons, P. (1995). Gender, ethnic and socio-economic differences in attainment and progress: a longitudinal analysis of student achievement over 9 years. *British Educational Research Journal, 21*(4), 465-485.

Samuel, J., & Bryant, P. (1984). Asking only one question in the conservation experiment. *Journal of Child Psychology and Psychiatry, 25*(2), 315-318.

Schunn, C. D., & Reder, L. M. (1998). Strategy adaptivity and individual differences. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation* (Vol. 38, pp. 115-154). New York: Academic Press.

Schunn, C. D., Reder, L. M., Nhouyvanisvong, A., Richards, D. R., & Stroffolino, P. J. (1997). To calculate or not calculate: A source activation confusion (SAC) model of problem-familiarity's role in strategy selection. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*(1), 3-29.

Secada, W. G. (1992). Race, ethnicity, social class, language and achievement in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 623-660). New York: Macmillan.

Shayer, M., Kuchemann, D. E., & Wylam, H. (1976). The distribution of Piagetian stages of thinking in British middle and secondary school children. *British Journal of Educational Psychology, 46*, 164-173.

Shayer, M., & Wylam, H. (1978). The distribution of Piagetian stages of thinking in British middle and secondary school children II: 14-16 year-olds and sex differentials. *British Journal of Educational Psychology, 48*, 62-70.

Siegler, R. S. (1987). The perils of averaging data over strategies - an example from children's addition. *Journal of Experimental Psychology-General, 116*(3), 250-264.

Siegler, R. S. (1988). Individual differences in strategy choices: Good students, not-so-good students and perfectionists. *Child Development, 59*(4), 833-851.

Siegler, R. S. (1995). How does change occur - a microgenetic study of number conservation. *Cognitive Psychology, 28*(3), 225-273.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Siegler, R. S., & Crowley, K. (1991). The microgenetic method - a direct means for studying cognitive-development. *American Psychologist, 46*(6), 606-620.

Siegler, R. S., & Shipley, C. (1995). Variation, selection and cognitive change. In T. Simon & G. Halford (Eds.), *Developing Cognitive Competence: New Approaches to Process Modelling*. Hillsdale, NJ: Erlbaum.

Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: Ho do children know what to do? In C. Sophian (Ed.), *The Origins of Cognitive Skills*. Hillsdale, NJ: Erlbaum.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology-General, 127*(4), 377-397.

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in Math and verbal self-concept, performance expectations, and motivation. *Sex Roles, 50*(3-4), 241-252.

Steffe, L. P., & Thompson, P. W. (2000). Interaction or intersubjectivity? A reply to Lerman. *Journal for Research in Mathematics Education, 31*(2), 191-209.

Stratman, J. F., & Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud protocols: Issues for research. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology*. Thousand Oaks, CA: Sage.

Tate, W. F. (1997). Race-ethnicity, SES, gender, and language proficiency trends in mathematics achievement: An update. *Journal for Research in Mathematics Education, 28*(6), 652-679.

Thelen, E., & Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, Mass.: MIT Press.

Thorndike, E. L. (1912). *Education: A first book*. New York: Macmillan.

Torbeyns, J., Verschaffel, L., & Ghesquiere, P. (2006). The development of children's adaptive expertise in the number domain 20 to 100. *Cognition and Instruction, 24*(4), 439-465.

van der Maas, H., & Molenaar, P. (1992). Stagewise cognitive development: An application of catastrophe theory. *Psychological Review, 99*(3), 395-417.

van Geert, P., & van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior and Development, 25*, 340-374.

van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. London: Academic Press.

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), 308-315.

von Glasersfeld, E. (1995). *Radical Constructivism: A way of Knowing and Learning*. London: Falmer Press.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250-270.

Vygotsky, L. (1978). *Mind in Society: The development of higher psychological processes*. Cambridge, Mass: Harvard University Press.

Wade, S. E. (1990). Using think alouds to assess comprehension. *The Reading Teacher, 43*(7), 442-451.

Wigfield, A., & Byrnes, J. P. (1999). Does math-fact retrieval explain sex differences in mathematical test performance? A commentary. *Contemporary Educational Psychology, 24*(3), 275-285.

Wilson, T. D. (1994). The proper protocol: Validity and completeness of verbal reports. *Psychological Science, 5*(5), 249-252.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89-100.

Young, K. A. (2005). Direct from the source: the value of 'think-aloud' data in understanding learning. *Journal of Educational Enquiry, 6*(1), 19-33.