Word Problems in Primary Mathematics: types of difficulties experienced by some "average" eight and nine year olds, and the effect of manipulating selected structural variables.

Elisabet Weedon, B.A. (Hons)

Thesis submitted for the degree of Doctor of Philosophy

Department of Education
Stirling University

1991

CONTENTS

		page
	Abstract	1
Chapter 1	General Outline	3
Chapter 2	Theoretical Background	18
Chapter 3	Research Methods and Approaches	67
Chapter 4	Exploratory Study	83
Chapter 5	Main Study: Phase 1	115
Chapter 6	Main Study: Phase 2	171
Chapter 7	Main Study: Phase 3	215
Chapter 8	Overview, Reflections and Conclusion	253
	Bibliography	
	Appendices	

ABSTRACT

This project investigates primary 4 children's difficulties when solving word problems. It consists of an exploratory study examining the feasibility of using task-based interviews in the school setting; and a main study divided into three phases. The tasks set to the children are selected/adapted word problems from SPMG textbook Stage 2.

Phase 1 investigates the difficulties of forty "average" primary 4 children from five different schools. Task-based interviews are used in conjunction with an error analysis.

Phase 2 makes structural alterations to six of the most difficult Phase 1 word problems to investigate more closely the possible cause of difficulty. These altered word problems are re-presented to the Phase 1 sample. The original problems are not re-presented to this sample as the task-based interviews allowed for considerable practice of these original problems.

Phase 3 took place a year later than Phase 2 and presents the the structurally altered word problems alongside the original problems to a different, but similar sample. This sample consists of 126 children from the five schools participating during Phase 1/2.

It is suggested that the findings do not support the view that a small unvarying number of variables consistently affect problem difficulty. Rather the sources of difficulty

are likely to stem from a number of highly complex interacting sources: and the language itself need not be the block it sometimes appears to be. Informal strategies were evidently important for a significant minority of children, particularly in relation to subtraction problems. This seems well worth investigating further. The use of these strategies suggested that the language of the word problem could be understood when the child could link it to his/her informal strategies. Also, given simpler numbers, the semantic implications of the problem could often be mastered.

CHAPTER 1 – GENERAL OUTLINE

1.1 Introduction and background to the project.

Aim of the project. The aim of this project is to investigate how primary 4 children cope with word problems in the classroom setting. Word problems from SPMG Stage 2 are used as this scheme is predominant throughout Scotland. A task-based interview is used initially to identify specific difficulties with selected problems. Structural alterations are then made to some of the word problems to explore these difficulties further.

Impetus for the project. The researcher's interest in children learning mathematics and the difficulties they seem to experience in doing so provided the impetus for this research project. An Open University degree which included the course "Developing Mathematical Thinking" together with a primary teaching course gave theoretical and practical insights into these difficulties. Classroom experience during teaching practices provided ample evidence of difficulties, particularly in relation to word problems.

It is interesting to note that the contemporary penchant for mathematics was not to be found amongst the ancient Greek. They considered it "a pure and lofty pursuit remote from the disorder of daily life, or, depending upon your point of view, an idle pastime having little usefulness" (Still, A. 1983, p. 302). Today, however, mathematics is an important subject. Yet long-standing concerns with standards within the subject seemed to have had little effect. An 1875 HMI comment (quoted by McIntosh, 1981) would be equally at home in today's press: "I must confess to some surprise at the extremely poor result in arithmetic" (p. 7). This perhaps points towards the need to improve the understanding of how children cope with mathematics in the classroom. It is hoped that this project will provide some useful insights into children's learning of one aspect of mathematics, albeit on a small scale.

This chapter will provide a context for those that follow by:

(i) considering briefly the nature of a word problem

(ii) looking at current trends within mathematics teaching through an examination of recent reports, and a discussion of current practice

(iii) considering the Scottish Primary Mathematics Group (SPMG) primary mathematics scheme as this provides the material to be used in this project

(iv) giving an overview of the research project

1.2 The nature of a word problem.

A problem according to Collins English Dictionary (Collins, 1978), is "a question or matter to be thought about or worked out". A word problem within mathematics refers to a question cloaked in language but which contains a mathematical operation or operations required to be worked out in order for an answer to be found.

Understanding of word problems. There are several stages, at least three can be easily identified, that the child has to go through in order to solve a word problem: the reading of the problem; the translational stage where the mathematical operation is extracted and converted into a mathematical representation; and the computational stage where the solution is attempted and hopefully achieved. It requires not only the ability to read but also the ability to read analytically to extract the information essential for computation, and to store it temporarily in a way that allows access when required for computation. This ability to read may be a specialised ability: mathematical texts differ from ordinary prose in a number of ways. It requires action to be taken in response to that which has been read; it is normally low in redundancy; and it requires mathematical symbols to be read and understood in addition to linguistic ones. These mathematical symbols are not necessarily read in a similar fashion to letters. Consider the meaning and reading/pronounciation of the symbol "2" in the following settings: 2, 0.2, 212, $\sqrt{2}$ and ½. Furthermore, there may be significant lexical differences between ordinary English and mathematical English

When presenting a child with a word problem the assumption is made that the child will read into the words the meaning that is intended by the author. Research into reading is now suggesting that the notion of a piece of writing as an entity with only one interpretation is outdated. The reader is seen as an interpreter of the text, and the interpretation put on the text will depend as much on the reader's background knowledge, experience and the situation as the actual writing in the text. To exemplify, using an example with specific relevance to mathematics consider this figure: 13 . It is likely to be interpreted as a B when in the following sequence: A 13 C , and as the number 13 in this sequence: 12, 13, 14 (Oatley, 1978, p. 46).

This is a brief overview of the type of features that may combine to make the reading and understanding of word problems difficult. Chapter 2 examines these features in greater depth and explores them in relation to past and current theories.

The importance of word problems in the curriculum.

Considering all these possible difficulties that may be encountered by the child when confronted with a word problem why are these types of problems considered an essential part of the contemporary mathematics curriculum? One reason may be that they provide an alternative to routine exercises for practising computational skills. Another reason for using them is to encourage children to apply their mathematical skills in different settings. Some would argue that the latter is the most important. If mathematics was not to be applied, it is perhaps in danger of becoming the "lofty pursuit" that it was considered to be in ancient Greece.

The skill of applying mathematics was emphasised by an HMI report from 1977 (quoted in Mathematics across the curriculum, 1980): "The only justification for including mathematics as part of the compulsory curriculum for all the children is the power it has to explain. ... But very few people obtain this power by learning mathematical skills in isolation ... Skills should be developed in a context". The HMI quoted above wanted the context broadened further into actual, rather than hypothetically practical tasks requiring the application of mathematical skills in a practical setting. One such practical setting could be a woodwork project requiring planning, measuring and actually producing a product. However, this type of practical mathematics rarely occurs within the primary classroom. If it were to be mathematically meaningful, the organisation required for such a project for a single teacher with a large number of children would be considerable. So, another way to present "realistic" problems has become word problems of this type: "Joe uses 4 nails to lay a floorboard. How many floorboards can he lay with 124 nails". This example comes from SPMG Stage 2 textbook (1985) and is used for practising division.

Bell (1980), amongst others, questions the usefulness of these "practical" verbal problems as an aid to learning to apply mathematics. He suggests that their routine nature does not demand any effort or creativity and that their solution can be learned through the application of certain routines. So, for example, "find the difference" indicates a subtraction problem and the smaller number must always be taken from the larger one. When this "routinisation" of the problem occurs the actual underlying structure of the problem can be bypassed. However, there is plenty of evidence, within this project, of children applying their own type of logic and solutions to these type of problems. These solution processes do not always reflect that which is encouraged by the teacher. So, perhaps word problems, at least for some children, provide a setting within which to practice mathematics creatively. Stephens (1977) argues not against word problems alone but against the trend of using more written material on an individualized basis in the classroom. The effect, he suggests, is to encourage an oversimplified language in order that it can easily be understood. Thus, it appears that the word problem can become, for some children, little more than a routine exercise that can be carried out with little understanding of the underlying structure of it. Verbal cues lead the child to the correct operation. It certainly is true that word problems within SPMG are used to practice specific arithmetic operations using the appropriate standard algorithm. The textbook page headings and advice given to teachers in the handbook (Teacher's Notes, 1985) do not seem to consider that children may have alternative ways of solving the problems.

The debate pursued by Bell (op. cit.) and Stephens (op. cit.) about the role or effectiveness of word problems will not be pursued any further in this thesis. The fact is that children do encounter them frequently in school mathematics. What is of interest here is how they cope rather than whether they should be solving such problems. Stephens' (op cit) comments on the increasing use of individualized materials in the mathematics classroom will be returned to in Section 1.3.2, where the focus will be on the current provision and practice.

1.3 Current Trends within Mathematics Teaching

1.3.1 Reports, recommendations and recent government proposals: a brief overview.

The Cockcroft Report. This has been by far the most important recent report on the teaching and learning of mathematics. It was published in 1982 and its remit was "To consider the teaching of mathematics in primary and secondary schools in England and Wales, with particular regard to the mathematics required in further and higher education, employment and adult life generally, and to make recommendations" (p. ix). It did not include Scotland; however, mathematics education in Scotland is considered to be similar enough for its comments to be taken seriously by mathematics educators north of the border. The report was prepared over a four year period. The report identified social, economic and technological changes in society over the last two decades. These, it suggested, have created a great deal of pressures on teachers. The need for improved training, both initial and in-service, was emphasised, as was support for the teacher at classroom level. The need for teachers to appreciate the great variation between pupils in rate of learning and level of attainment was pointed out. The need for of a varied mathematical diet was also raised. If pupils are to develop the skills to apply their mathematics in new settings practice of computational skills alone is not sufficient.

In addition to the Cockcroft Report a series of three reviews on research commissioned by the Cockcroft report have been published. Most relevant to this research has been the review by Bell et al (1983) entitled "Research on Learning and Teaching". It looks in detail at the stages of cognitive development and its importance to mathematics education. It also comments on the great increase in the use of written material in the classroom and the demands this places on the pupils. Thus both these reports have raised questions in relation to the topic of this research.

Earlier reports, going back to the end of the last century all contain recommendations similar to the Cockcroft Report. Progressive methods in education have been accused of lowering standards. However, it is questionable if these methods have in fact ever been fully implemented. The Cockcroft report seems in some places to echo recommendations made long ago. Contrast the following quote from the Cockcroft report "The learning of number facts ... needs to be based on understanding" (p. 87) with ".. no instruction in the rules of arithmetic can be really valuable unless the process has been made visible ..." (Reports 1895, McIntosh, op. cit. p. 9).

Thus recommendations on how mathematics should be taught have been available to teachers for a long period of time. It seems, however, that any changes that have been made have not had the desired effect - that of producing adults and children who can use mathematics effectively.

1.3.2 Current Practice

Changes within the teaching and learning of mathematics. These are commented on by Bell et al (op. cit.) in "Research on Learning and Teaching". They state that within the last twenty years the increasing demand for pupils to read and write in the mathematics classroom has affected the language used in textbooks and on workcards. Textbooks used to provide examples to be worked through in a teacher led lesson for the whole class. Nowadays pupils are expected to work through workcards and textbooks individually. This change stems from a change in the perception of children's learning. It has influenced educational policy. Earlier theories of learning proposed by behaviourism emphasised the learner as an empty vessel to be filled with knowledge. Consider the following statement from one the most well known behaviourists B.F. Skinner: "The school is concerned with imparting to the child a large number of responses of a special sort" (1975, p. 158). He then goes on to discuss how the child's behaviour can be shaped and reinforced so that only the correct responses are retained by the child. Contrast this view of the learning child with a comment from von Glaserfeld "Children, we must never forget, are not repositories for adult 'knowledge' but organisms that, like all of us, are constantly trying to make sense of, to understand their experience" (1987, p. 12).

This idea that children or indeed all human beings process and interpret information from the environment has already been encountered when discussing the process of reading. It is discussed at length in Chapter 2. Piaget has perhaps been one of the most influential theorists in early mathematics. He suggested that knowledge about the world is stored internally as "schemata". These schemata are gradually built up and elaborated as the child actively interacts with the world around him/her. The schemata develop at different rates for different children depending both on genetic and environmental factors. The Piagetian influence has led to a greater emphasis on the individual child's need to learn at his/her own rate. Piaget's insistence on the need for active involvement has also had an effect on contemporary mathematics schemes. Much more emphasis is now put on practical activity in the classroom.

These two taken together: the individual's unique rate of development, and the need for practical activity has probably led to an increase in ability groupings within the primary classroom. Instead of the whole class being taught mathematics at the same time, only maybe a third of the class may actually be involved with the teacher being explicitly taught mathematics. The rest of the class may be involved in mathematics but using workcards, workbooks or a textbook, or doing another subject. The teacher, in this situation, has to cater not only for one group but for three or more groups with different needs. Within these groups the range of abilities will also vary. Some need high level work to stretch their abilities others require help with the simplest type of mathematics. The Cockcroft report (op.cit.) speaks of the "seven-year gap", suggesting that within any one class of eleven-year olds, the gap between the most able and the least able can be as much as seven years. The gap may not be as great with younger children. However, the following extremes were found within one class in this project: one child who scored the highest in the standardised mathematics test out of all five schools; and a child who was unable to name "the number before" a number below fifty. This child had a number line in front of him. This wide range of abilities puts a great deal of pressure on the teacher when it comes to catering for individual needs, especially as there are few primary teachers who are mathematics specialists. The "complete package" mathematics scheme provides an answer. The individual child can then work to some extent at his/her own speed. However, to become skilled in all aspects of mathematics require not only the development of basic arithmetic procedural skills but the development of higher order skills.

A more recent study by Desforges & Cockburn (1987) suggest that the development of these higher order skills is very difficult, even for the competent teacher, in the ordinary classroom setting. It is further suggested that experimental evidence shows that very little is known and understood by researchers on how to encourage the development of these types of skills. Thus to expect teachers to be able to develop these skills in large numbers of children who are at different stages level of development is perhaps to ask too much of the individual teacher. What also needs to be considered here is that the primary teacher has to cover a wide curriculum, of which mathematics is only a part. It is understandable if the teacher feels that a commercially produced scheme may provide a better mathematics education than an education that relies on the teacher's own ideas. Perhaps the combination of individualisation of learning and the increasing breadth of the curriculum has led to the growth of commercially produced schemes within the primary classroom.

It has been suggested in this section that a change in the view of how children learn has had an effect on the organisation of the classroom and resulted in an increasing reliance on written material to cater for differing individual abilities.

1.4 Description of and reason for use of SPMG within this project

Description of SPMG. SPMG is probably the most widely used mathematics scheme within Scotland and it claims to be the most widely used British scheme. The scheme originated in a working party on mathematics in the primary school (National Primary Mathematics Project). The working party's remit was to examine the structure and content of mathematics in Scottish primary schools. Constant consultation between the authors of the scheme and teachers has led to a revised edition of the scheme. This later edition takes into account comments from teachers and changes within the subject itself.

For example, one such change has led to the introduction of calculator activities in SPMG

Its full title is "Primary Mathematics: A development through activity". Its claims amongst others are to provide a proven structure, material for a wide ability range and strong support for the teacher (Heienemann Educational, 1989). The materials included in the scheme for each year group are: workbooks, workcards, a textbook, teacher's notes, separate answer book, teacher's materials pack and progress tests. A varied use of graphics and colour is in evidence in most of the material. Of these the most commonly used by the pupils are the workbooks, the textbook and workcards. The teacher's notes emphasises the progressive nature of the material and also the fact that the practical activities are essential for the development of understanding within mathematics.

The workbooks and the textbook are seen as the basic essentials of the scheme. The teacher's notes are considered of importance as they contain many suggestions for practical activities and oral work and games which will help to introduce the work to the children. It is claimed that the "range and extent of the work in any one Stage is such that a child of average ability may be expected to complete it in a school year. It is not intended, however, that every child will complete every page and card ..." (Teacher's Notes, p. 6).

The reason for using SPMG within this project.
SPMG has been chosen to provide the material for this project because it forms the main scheme within Scottish schools. It is not the aim of the project to prove, or disprove, its effectiveness. Rather what is intended is to consider how children actually cope with some aspects of it.

1.5 Overview of the research project

The introduction stated that the aim of this research was to investigate difficulties experienced by Primary 4 children when solving word problems. To explore how the individual does deal with the written problems without any outside help, a small number of word problems have been chosen from the textbook to form the core of the research material. How the children interact with this material and possible permutations of it are examined. The project aims to look at the processes children use when tackling word problems likely to be encountered in the classroom setting. Not only correct responses are of interest but the strategies used and the type of errors produced by the children are analysed. It is hoped that this type of analysis will throw more light on how children cope with the problems they have to solve.

In the following chapters:

Chapter 2. examines the theoretical background to the research and provides a context within which the findings can be discussed.

Chapter 3. discusses research methods, particularly those chosen for this project.

Chapter 4. explains the exploratory study where the feasibility of using particular methods in the classroom setting is explored.

Chapter 5. forms Phase 1 of the main study. It explores a number of "average" children's difficulties with word problems through the use of the task-based interview.

Chapter 6. forms Phase 2 of the main study. It returned to the children from Phase 1 with a number of rewritten problems, to explore the effects of changes made to them. These rewritten problems were based on the difficulties evidenced through Phase 1.

Chapter 7. discusses Phase 3 of the main study. A new sample selected, similar to Phase 1/2 but covering the whole ability range. This provided a less time-consuming, but equally effective data gathering method for this part of the project. The rewritten and original problems from Phase 1/2 were used.

Chapter 8. sums up the results from the project and suggests further research.

CHAPTER 2 - THEORETICAL BACKGROUND TO CHILDREN'S UNDERSTANDING OF WORD PROBLEMS.

2.1 Introduction

This research project has set out to look at how Primary 4 children cope with some of the word problems in S.P.M.G. textbook Stage II. The aim has been to assess problem difficulty based on the children's ability to solve a number of particular problems. Some aspects of the problem difficulty were then examined in greater depth. However, word problems in themselves do not provide the only guidance to their difficulty; it is in interaction with the problem solver that the difficulty arises. Thus it is necessary in this chapter - the theoretical background to the research - not only to look at the structure of the word problem that may cause difficulties but also what the child - the problem solver - brings to the problem, and the aspects outwith the problem that may affect problem difficulty. Within this latter category would be considered such matters as page layout and illustrations. Three broad strands can thus be identified for this chapter:

(i) the word problem itself.

(ii) the presentation of the word problem.

(iii) the child and his/her cognitive development and how this development affects the developing understanding of mathematics, particularly in relation to word problems.

This leaves out one important aspect in the child's understanding of word problems:

(iv) the teacher variables. Baroody & Ginsburg (1986) discuss "the effects of schooling". This research does not aim to explore these types of variables - however, they undoubtedly do have an effect. A brief discussion on research relevant to this area will therefore be included in this chapter.

Accordingly this chapter will take the following format:

(i) An examination of theories that attempt to account for learning with particular reference to the development of mathematical knowledge. Two theories will be discussed: behaviourist and schema theory, along with their differing attitudes to the role of memory. The different types of knowledge regarded as contributing to mathematical understanding will also be examined here.

(ii) An analysis of the research relating to word problems. This will include looking at the structural aspects internal to the problem and the specific effect of the language used in word problems.

(iii) A discussion of the effect of page layout and illustrations. This area has not received as much attention from researchers as have the two preceding areas; this section will therefore be brief.

(iv) As mentioned, "the effects of schooling" have not been focused upon in this research project. However, they are important, and research from this area should be considered briefly.

(v) In conclusion, the various strands will be brought together and possible interactions will be discussed. The relationship of this chapter to the rest of the research will conclude this chapter.

2.2 The cognitive development of the child and the development of mathematical understanding

2.2.1 General theories of learning.

Two main, opposing theories will be considered here: behaviourism/learning theory and schema theory. These two theories are discussed here because they have both had an effect on the instruction and organisation of mathematics teaching, and these effects can still be identified within current educational practice.

Behaviourism was the dominant psychological theory during the early parts of this century up until the late 50's and in revised forms it is still influential. It has had a considerable effect on the organisation of teaching and therefore merits a discussion here. It suggests that the individual is totally created by the environment with very little effect of inherited genetic abilities. All behaviour is governed by stimuli which produce certain responses. For a behaviour to become established an association needs to be formed between the stimulus and the response. The most effective way of creating an association is by positively reinforcing the required response. Pavlov's dog is an example of this. Once the dog had learnt that a bell was connected with food presentation it learnt to salivate when it heard the bell. Learning mathematics is explained in similar terms - once the child has learnt to associate 2 + 2 with 4 s/he will respond automatically with this response, provided enough reinforcement is provided, and that incorrect responses are discouraged. Teaching thus becomes simply a management of stimulus and reinforcement. Skinner (op.cit.), one of the main proponents of behaviourism believed that if the correct stimulus and reinforcement could be worked out more effective learning would automatically ensue. Learning had to be built up successively, in small stages with appropriate reinforcement. Programmed learning is based on this idea. Rote learning and drill also form part of this approach, and retain a place in much of today's mathematics teaching. Good practice in mathematics teaching became a matter of building up strong associations between the number facts that had to be acquired and the desired responses.

The method used for investigating learning was strictly limited to behaviour that could be observed objectively (Morris, 1974). The role of mental processes was not studied because it could only be inferred. The emphasis on objectivity led to the development of psychometric tests of a standardised format which has guided much of the test construction and examination procedures until recently. It has one great problem - that of finding exactly what acts as a stimulus in a world that is buzzing with stimuli. It is easy, in the case of Pavlov's dog to keep the dog in a highly controlled environment; this cannot be done to humans except perhaps in unnatural environments such as prisons. Also by limiting itself to studying the observable it is in danger of limiting its domain to the trivial. Laboratory induced behaviour does not necessarily reflect behaviour outwith the laboratory. Much of its research has been carried out on animals (see for example Thorndike's research as reported in Resnick & Ford, 1984) - but people do not necessarily follow the same behaviour patterns as animals.

This approach was more dominant in the United States than in Europe. It was being questioned as early as the 1930's in Britain when Bartlett (quoted in Baddeley, 1976) suggested that memory was organised in a meaningful manner. Bartlett argued that central to learning was the individual's "effort after meaning". In other words, each individual interpreted the environment in a manner meaningful to himself or herself. Bartlett used meaningful texts rather than nonsense syllables to study memory and found his subjects seemed to remember the texts in a manner that fitted in with their previous knowledge. To explain this phenomenon Bartlett borrowed the term "schema" from neurology research. In neurology it was used to explain the individual's knowledge of the relative position of his or her limbs. Bartlett used it to suggest that individuals create models or schemata in their minds based on their experiences of how the world around them functions. It is interesting to note a more contemporary definition of schema: according to Anderson (1984, p. 5) it is "an abstract structure of information. It is abstract in the sense that it summarizes information about many particular cases. A schema is structured in the sense that it represents the relationships among the components". It thus links in with Bartlett's idea in that it is a structured abstraction of the reality that provides a model for the individual and governs his/her response to all aspects of the environment.

Another early proponent of schema theory is Piaget (see Donaldson, 1978 for an overview of Piagetian theory). His studies focused on children as they develop and he used the concepts of accommodation and assimilation to explain the growth and elaboration of schemata. Assimilation occurs when a new piece of knowledge is fitted into an existing schema; accommodation refers to the reorganisation of a schema in order to allow the new knowledge to fit in. As the schemata are created through the child's interaction with the environment Piaget emphasises the importance of practical activity in learning. The young child learns about the world not only by observing and listening but through doing - practical activities through play provides the child with early learning. Piaget also suggested that the child develops through a number of invariant stages. However, the evidence for these stages is now being questioned and cannot be considered a certainty (see e.g. Groen & Kieran, 1983). Another area that has been questioned is his reliance on language to examine evidence for competences in children. Donaldson (op.cit.) suggests that the results in some of the classic conservation tasks are due to the way the questions are phrased rather than the inability in the children to carry out the task. Her research supports this contention and many other researchers have found similar results (e.g. Bryant, 1975 and Smedslund, 1979). Whilst some aspects of Piaget's theory are now being questioned his legacy to developmental psychology is considerable. He pioneered methods of observing and interviewing children that are still being used and he has left a wealth of data and ideas. The fact that in his later years he modified a number of earlier ideas is important - he recognised the need to change and rebuild his theory as new evidence appeared.

Bartlett and Piaget have been followed by others questioning the behaviourist approach. There is a considerable body of research that does not accept that the young child is an empty vessel waiting to be filled with knowledge. Piaget's theory is a general, global theory of learning; more specific theories have been put forward dealing with more specific aspects of learning but still drawing on the general idea of schema theory. Language learning is one such area. It will be examined briefly here for two reasons: firstly, this research looks at word problems which involves the use of language, and secondly because some of the theories within mathematics have drawn on these earlier language theories.

Many of these theories draw on the cognitive psychology approach known as information-processing. A number of these theories have used computers in order to simulate possible human behaviour. The development of computers have had a considerable impact on the type of theories developed by cognitive psychologists. The development of computer simulation has the advantage that they have to be highly specific in order to work. The main disadvantage is probably that they lack the flexibility of human behaviour. This point will be returned to when discussing specific computer models that attempt to explain arithmetic knowledge.

2.2.2 Schema language theories.

A number of slightly different, but essentially similar, theories have been proposed. The most commonly used terms are schema or schemata but Minsky (1977) uses the term "frames" and Schank & Abelson (1977) refer to "scripts". These see knowledge as packaged networks of concepts and information based on each individual's repeated experiences. Individuals store their knowledge and experiences in these working packages, and select them from storage as appropriate to deal with incoming experiences and information. This means that the individual who possesses these schemata, scripts or frames is an active participant in their use, elaboration and application. For example. the listener/reader interprets the incoming information according to his/her existing, selected schemata. So, the sentence "Mary had a little lamb" can invoke different schemata: in the context of a meal it is in relation to Mary's portion of food; in the context of telling stories to children it is the title of a nursery rhyme. Thus the selection of a relevant schema depends on the individual's interpretation of the situation. Communication through language depends on the individuals involved in the situation to have sufficiently similar schemata. Context often provides additional cues to help the selection of a reasonable schema. For example, the phrase "it looks like rain" could be interpreted in a number of ways depending on the situation. It may just be a polite way of passing the time when meeting someone in the street, it may be a clue to get the washing in, or it could be interpreted as a suggestion to put on a waterproof jacket before venturing out. In all these instances, few adults would have any difficulty in interpreting this phrase. However, this may not be the case for younger children. Research has shown that children may not attend to the same cues as adults. Consider the following sentence: "The Smiths saw the Rocky Mountains flying to California". Adults would correctly infer that the Smiths flew in an airplane to California and whilst so doing saw the Rocky Mountains. Til Wykes (quoted in Johnson-Laird & Wason, 1977) found that four-year old children did not interpret this sentence in the same way - they were more likely to think it was the Rocky Mountains flying. Most parents can probably cite similar instances when young children have interpreted adult statements in their own, usually logical, but non-adult fashion. As children grow up their ability to interpret statements in a more adult fashion increases. However, at what stage the child reaches an adult type of understanding has not been ascertained , and it is likely to vary between individuals. This aspect of language development is mentioned here, because it is considered to be of importance when looking at word problems. What seems to be a very simple, unambiguous statement to an adult may not be so for a child.

What has been suggested here, in this brief look at language theories based on schemata, is that these are internal structures based on the individual's past experiences. As experiences vary from person to person there are bound to be a number of individual differences. However, shared cultures allow for the development a schemata that are sufficiently similar for communication to take place. It is also suggested that as children have not had as many experiences as adults they are likely to be more limited in the ways that they can interpret situations. This is perhaps best summed up by Bransford & McCarrell's (1977) statement that the information contained in a sentence is "depending on the cognitive contributions that the comprehender makes" (p.386). Whilst language learning has been focused upon it is suggested that schemata govern all human behaviour and interaction.

2.2.3 The role of memory in learning mathematics.

The very word memory conjures up different feelings in different people. The emphasis placed upon rote learning and memorization has led some people to feel that memory should not be overemphasised in the learning process. However, the use of the word memory here is not intended in that manner;

rather it is used in the broadest sense, and it stands for the brain's way of organising and retrieving information about the world around it. What is being discussed is how this external knowledge is absorbed into an internal structure of a kind, how it is stored and retrieved, thus the whole system of knowledge organisation and retrieval is under consideration. As can be seen in Section 2.2.1 the early behaviourists did not consider memory as part of their sphere of study as it could not be directly observed. However, later researchers within this field found it impossible to progress without suggesting some sort of model of internal brain functions and attempts were made to explain the memory structure as a series of internal stimulus - response chains. Developments within computers and information processing then provided ways of modelling the memory functions. Memory was seen by many (Byers & Erlwanger, 1985) as divided into three compartments: long-term memory which is of unlimited capacity and provides permanent storage; short-term memory which retains information for short durations and selects information for storage in long-term memory; and working memory which stores information essential for constructing suitable representations of the information. This last memory component would be essential in mathematics for constructing a suitable representation of the problem to be solved. Long-term memory storage, it is suggested, is in terms of networks and based on semantic content of the information. Piaget and Bartlett and those that have become known as constructivists suggests storage of information in terms of inter-related semantically based schemata. It may be suggested that the terms semantic networks and schemata are just two different terms for an essentially similar structure. However, the crucial difference between the two approaches is probably expressed by the term "constructivist". It implies that the learner constructs his/her own schemata and interprets incoming information in light of what has already been stored. This point is emphasised by Ginsburg (quoted in Byers & Erlwanger, op.cit.) when considering mathematical difficulties. Ginsburg suggests that "errors are seldom capricious or random", rather they conform to the logic available to the child in the problem solving situation. He further suggests that these errors are often based on a set of erroneous rules that the child has created from a related piece of learning. So, for example, the child that states the 3 x 4 is 7 may have misread the multiplication symbol as an addition symbol for a variety of reasons. The response 7 shows that the error is not random rather it is "misapplied" knowledge. Van Lehn (1983) would also support this statement: mistakes made when using the standard subtraction algorithm are usually based on a correct piece of knowledge applied incorrectly. Take the sum:

$$\begin{array}{r} 345 \\ -158 \\ \hline \end{array}$$

a common response to this sum is 213; this Van Lehn's suggests is due to the student having learnt, correctly, that you cannot take a larger number away from a smaller number (until you start learning about negative number). Again evidence that children's behaviour is not random but likely to be based on misapplication of a rule.

So, it can be seen that a model that considers the function of memory is important when considering how mathematics is learnt. The knowledge necessary for learning and using mathematics needs to be stored and accessed so that it is available when required. Whether this structure should be called memory or given another name could be debated. Neisser (quoted in Byers & Erlwanger, op.cit.) suggests it should not be retained and this section will be concluded with his exhortation "What we want to know ..... is how people use their own past experiences in meeting the present and the future" (p. 278).

2.2.4 Schema theories in mathematics.

The concept of schema development as an explanation of how mathematical understanding is created and grows has been used by many researchers. Two different approaches can perhaps be discerned - those that use computer simulation methods to test theory, and those that use interviewing as their main, but not necessarily only, method. Those using computer simulation tend to limit their theory to specific types of mathematical understanding, for example, by trying to account for the development of addition within the early school years. Those using interviewing type techniques tend to look at mathematical understanding from a broader perspective. It is probably fair to say that both are useful and perhaps necessary. Without the precision provided by the computer simulation type of experiments theories may not be well thought out, but without the broader outlook mathematics education may become restricted to examining the type of knowledge that can easily be represented in a computer simulation programme.

Broad perspective schema theory.

Amongst those looking at the broader outlook are Baroody & Ginsburg (op.cit.). They refer to schema theory as "the alternative theory". In their views it is an alternative to those theories that still maintain a basically behaviourist approach. Siegler (1987), Campbell (1987) and Graham (1987) all suggest that mathematical knowledge is stored in networks but that association, built up through practice, between problems and answers accounts for mathematical knowledge. Although they suggest a network storage, and thus making inferences about internal mental operations, the behaviourist contention that stimulus-response chains are the building blocks of learning is in evidence. Baroody & Ginsburg argue against association as accounting for mathematical learning. They suggest instead that arithmetic knowledge is stored as related representations of basic number facts. This makes for a more economic storage of number facts. If 6+2=8 is known, then 2+6=8 will also be known. Associationists would suggest separate storage for these basic number facts. The storage is based on meaningful relations between the items and is thus considered a semantic model. Their argument is based on evidence that, once certain underlying rules are understood, answers can be generated for previously unseen problems. For example, if a child has learnt that $N - 0 = N$ through a number of specific instances they can give this response to a previously unseen sum. Associationist theories do not seem to be able to explain this phenomenon. This argument has also been presented to support schema theory within language development. Once a certain level of language development has been reached previously unseen sentences can be understood.

Anderson's definition of a schema as an abstract structure which organises information is quoted on page 23. Anderson goes on to suggest two different forms of schema: strong and weak. Strong schema is "principle driven" and predictions made based on this type of schema are derived from these principles. Weak schema is "precedent driven". Here it is suggested decisions are based on evidence which is "looked up". This latter type of schema, Anderson proposes, is the type of schema that accounts for most behaviour. Baroody & Ginsburg (op.cit.) use this idea of weak schema to explain much mathematical behaviour. Only when thorough understanding is achieved can a strong mathematical schemata be created. Because so much teaching is based on mathematical routines and rules weak schemata are developed. Instead of being able to apply a principle in order to find a solution, precedents are looked for and these guide the solution. This has the effect of limiting the type of problems that can be solved to ones that are perceived as sufficiently similar to a preceding one that has been successfully completed.

Support for a semantic model of mathematics also comes from Davis (1984). He suggests that theory is vital in understanding mathematical education and also that it is important to go beyond the observable. Support for this comes from the physical sciences - the atom, for example, cannot be seen but yet its existence is accepted. He uses the concept of "frames" to explain how mathematical knowledge is acquired and stored. Davis uses the term "frames" instead of schemata, drawing on Minsky's work in language (see Section 2.2.2 p. 25). The structures he is considering are like schemata, so the two terms are seen as interchangeable. He suggests that there are a number of commonly shared frames to be found in elementary arithmetic which are similar from individual to individual and also between different cultures. From these basic type frames develop further frames some of which are likely to be idiosyncratic to fewer individuals.

Specific theories: Part/whole theory.

This is a specific model of the early development of addition and subtraction understanding, based on schema theory. A detailed account of it is given in Resnick (1983). The term part/whole is used to denote the relationship between quantities. For example, if the number 7 is taken to be the whole, 2 and 5 can be seen as part of this whole. Given 2 and 5, the whole - 7 - can be found; given 5 (or 2) and 7; the part 2 (or 5) can be found. Prior to development of part/whole schemata the child is considered to have a number line schema which does not have access to this type of understanding. The number line schema allows only for understanding of number in terms of before and after. In relation to the numbers 2, 5 and 7, the child can only say that 7 is 2 more than 5 by counting back 2 from 7 and discovering that this leads to number 5. In other words, numbers are not seen as consisting of other interrelated numbers. This model of development is supported by research (Carpenter, Moser & Bebout, 1988) which shows that young children can calculate simple arithmetic problems through direct modelling, that is by making an external model by using blocks or other items to represent the problem statement. If direct modelling is not possible due to the wording of the problem statement they cannot solve the problem. An example of a problem that allows for direct modelling would be "David had 5 apples, Alison gave him 3 more apples. How many apples does David have now?" This would allow the child to put out five blocks to represent the initial quantity, put out another three to represent those added and count the total quantity. An example of a problem not allowing direct modelling would be "David had some apples, Alison gave him 3 more. David now has 7 apples. How many did he have at the beginning?" Because there is no exact quantity to represent initially the child with only number line understanding cannot directly model this problem and cannot therefore solve the problem. Thus early mathematical understanding is limited as it allows only for relating quantities as being either smaller or larger. A major development occurs (probably in early school years) when the part/whole schemata start to develop. These are an elaboration of the earlier, more limited schemata. These part/whole schemata, which according to Resnick are limited in number, develop throughout childhood and possibly throughout adulthood into an elaborate network.

Resnick based her theory on the work of Groen & Parkman. This work looks at possible strategies for solving simple mental addition and subtraction problems based on reaction time measurement. The longer the time the more cumbersome the strategy. The simplest strategy for addition involves mental counting of all the numbers involved, for example, 5 + 3 would be done by first counting up to five and then carrying one with three more up to eight. This is known as the "sum" model Here, according to the theory the child would be using a number line schema. The most sophisticated strategy would start by setting the mental counter to the highest number, irrespective of its position in the number sentence, and then increment this by the smaller number. This is known as the "min" model as the reaction time would be based on the minimum addend. This, the researcher suggests is evidence of the emergence of part/whole schema, because it shows at least an implicit understanding of commutativity. These reaction time studies were supplemented with individual interviews with children and compared to adult reaction times. A number of different accounts such as a pair/equivalence and a default model as well as the part/whole model were considered to explain the results of Groen & Parkman's research. Pair-equivalence depends on practice of specific pairs in order to create an association between these. In the Groen & Parkman experiments the most efficient strategy - the "min" strategy emerged without such practice. Thus the pair/equivalence model was discounted by Resnick. The default account suggests that the child would recognise and come to utilise the order invariance principle through practice of counting objects. This would lead eventually to the adoption of the "min" strategy. The default model is not completely discounted by Resnick, but it is suggested that part/whole theory provides the most economical account of the development of this strategy and other similar strategies. The part/whole schema puts the three numbers into a complementary relationship. This theory is thus considered as the most plausible.

Specific theories: Computer simulation models.

These types of models have been developed by Resnick (op.cit.), Riley (quoted in Riley, Greeno & Heller, 1983) and Briars & Larkin (quoted in Carpenter, 1986) They are based on part/whole schemata analysis and despite minor differences are quite similar. As mentioned at the beginning of Section 2.2.4 these models generally attempt to explain only a limited area of mathematical behaviour. To demonstrate the type of computer simulation models that have been developed to account for early arithmetic understanding one of these will be examined here - that of Riley (op.cit.). She has used evidence from research using interviewing techniques to study young children's understanding of arithmetic word problems. From this evidence the type of knowledge required to solve different kinds of problems has been suggested. Three different types of schemata have been suggested as necessary: a problem schema which is used to create a suitable representation of the problem; action schema, that contains knowledge about actions used in planning solutions to problems; and strategic knowledge which is used for planning solutions to problems. These types of schemata increase in number and complexity as the child matures and learns more complex mathematics. Prior to these types of schemata being in existence the child relies on a schema that is akin to a number line. This type of schema allows only for quantities to be calculated according to a "larger than" or "smaller than" principle. Three levels of mathematical understanding have been identified using the model that children move from the simpler number line type of schema to the differentiated schemata that imply a part/whole understanding. At level 1 the child is limited to problems that can be externally and directly represented (for example, by using wooden blocks), but according to Riley, the child still applies an internal schema to the problem solution. Not all researchers agree with this interpretation and this will be discussed at the end of this section. Level 2 forms an intermediate level. Here the child can maintain an internal representation of changes made to external objects. At level 3 the development of part/whole schemata allows for internal representation and manipulation of symbols, without the use of external objects.

Riley suggests from research evidence that it is often a difficulty in understanding and representing the information that causes mathematical difficulty. According to this view the difficulty occurs in the problem schema rather than the action schema. She cites as evidence Hudson (quoted in Riley, et al, op.cit.) where a change in wording in the question asked of the children produced a dramatic increase in the children's ability to carry out the task.

A different type of computer simulation model has been produced by Van Lehn (1983). It looks at the type of errors that students produce when using the standard subtraction algorithm. It differentiates between "errors" and "slips". Errors are non-random mistakes usually based on a faulty application of a procedure that is correct under certain circumstances. Slips are just mistakes occurring because of lack of attention. It is the errors, or "bugs" as Van Lehn calls them, that are central to his simulation models. In a sense he is looking at a breakdown in what Riley would call the action schemata. Maybe this is due to the different age groups that the two researchers are working with. Van Lehn's older students are operating within a setting where understanding of formal algorithms have become important. When these have not been properly understood error occurs. Riley's subjects were all pre-school or early school years and still relied to an extent on informal understanding of mathematics. Van Lehn's research has built on and extended that of Brown and Burton. It has looked at student errors and created a theory that discusses and explains the nature and pervasiveness of errors that occur when students use the standard subtraction algorithm. It is called repair theory. It suggests that the student learns, through instruction, a number of core procedures for carrying out mathematical calculations. When these core procedures are incomplete, through, for example, missed lessons or lack of time for assimilation, bugs are created. The student, according to this theory, will then try to "repair" the core procedure by inserting an essentially correct piece of information in the wrong place. These bugs show up as errors in the calculation.

Both these researchers have created models that account for mathematical behaviour within a small area of mathematical understanding. They provide a precise method of examining possible explanations of children's mathematical understanding. However, as Carpenter (op.cit.) states, they do not account for all the complexity of children's behaviour. It is probably fair to say that the requirements of a computer simulation model that is to function cannot take into account idiosyncratic behaviour. Thus when creating these models order has to be imposed on the data and important aspects of behaviour may be left out. Langford (1986) also questions the interpretation which has been put on the data by Riley. He suggests that instead of postulating internal representation of the data it can be equally well explained by positing an entirely external representation. The children involved in these studies used blocks to represent the sums and thus Langford argues had no need to create an internal representation. The data that has been explained by these theories could perhaps be seen as those which Davis (see Section 2.2.4 page 33) refers to as "commonly-shared frames".

2.2.5 Types of knowledge in mathematics.

The simulation model of Riley presents a suitable starting point for this section. She distinguished between problem and action schemata and these types of schemata were based on different types of knowledge. Problem schemata relate to the underlying conceptual understanding required for interpreting the problem, action schemata refer to the procedural understanding needed for executing the solution once the problem has been understood and suitably represented. Thus the two different types of knowledge: conceptual and procedural, that are considered to underpin mathematical knowledge. Hiebert & Lefevre (1986) examine these two types of knowledge and conclude that despite differences in nomenclature these types of knowledge have been identified in several theories. So, for example, Piaget uses the words "conceptual understanding" and "doing" and Anderson distinguishes between "declarative" and "procedural" knowledge. Yet another mathematics researcher, Skemp (quoted in Silver, 1986), talks of "relational" and "instrumental" understanding. What is in essence being discussed is the relationship between understanding the concepts that underlie an action and actually carrying out that action. In mathematics this would refer to understanding why a particular action is correct as well as being able to carry it out. Hiebert refers to the understanding as conceptual knowledge and the ability to carry it out as procedural knowledge and these two terms will be used here.

Hiebert & Lefevre (op.cit.) distinguishes between the two types of knowledge in the following way: conceptual knowledge "is characterized most clearly as knowledge that is rich in relationships" (p. 3). It develops through an increase in the links between items of knowledge. It can be accessed at random. In contrast procedural learning tends to be of a serial nature, one item of knowledge has to be retrieved and acted upon before the next is called up. So, for example, in the standard addition algorithm if the units column contains two digits with a value greater than ten this "ten" has to be carried into the tens unit before the tens can be added correctly. Procedural learning in mathematics, according to Hiebert & Lefevre (op.cit.) consists of two parts: the formal symbol representation system that makes up mathematics; and the rules or algorithms used to manipulate this system.

There exists within mathematics education a debate which has gone on for decades and is still continuing as to which type of knowledge is the most important for learning mathematics. At one extreme, there are those that maintain, in the tradition of the behaviourist type theories, that if the procedures are learnt correctly then the child is capable of doing mathematics. In contrast to this position others suggest that procedural learning is useless without a thorough understanding - perhaps Bruner (see discussion in Orton, 1987, pp. 83 ff.) could be considered to hold this kind of position. Within these two extremes there are varying views putting different emphasis on the two types of knowledge. It is probably wise to consider a position in between the two extremes: the formal symbol system and associated procedures provide a very powerful tool for carrying out mathematical calculations. It has been created over centuries. It is difficult to imagine a discovery learning situation where all this could be discovered by a child. This type of knowledge falls into the category of what Van Lehn (1983) refers to as "non-natural" knowledge. He makes a distinction between this, and the other type of knowledge which he calls natural - this is learning picked up by the growing child without formal tuition. Spoken language would fall into this category. Written language would not, it is normally learnt through direct tuition, hence it is non-natural. A parallel can perhaps be drawn in mathematics: counting could be considered as natural learning. Many researchers (e.g. Carpenter, Hiebert & Moser, 1981, Gelman & Gallistel, 1978) have shown the pre-school child to be accomplished in these tasks. However, using the formal number system (the "writing" of mathematics) is usually learnt through direct tuition. Thus it would seem necessary for the child to become competent in both types of knowledge in order to become an effective mathematician.

There is a further difficulty when considering these types of knowledge in relation to learning mathematics. Silver (op.cit.) points out that it is extremely difficult to isolate the two types of knowledge as they are usually inextricably linked - in order to demonstrate conceptual understanding procedural knowledge is used. For example, if a child can identify an equilateral triangle s/he may just have learnt a set of rules to determine whether a shape comes into this category or s/he may have deeper conceptual understanding. Silver therefore suggests that what should be studied is the interrelationship of these types of knowledge.

Another question emerges when considering the inter-relationship of conceptual and procedural knowledge: that of which comes first. Is procedural knowledge based on conceptual knowledge? In other words, is it impossible to learn a procedure before conceptual understanding has been achieved. Here, research opinion diverges yet again. According to Carpenter (op.cit.) the models of Riley and that of Briars & Larkin would support precedence for conceptual understanding. This is questioned by Carpenter who comments that the models involved have oversimplified the children's problem solving behaviour and thereby placed undue emphasis on conceptual knowledge. Baroody & Ginsburg (op.cit.) also disagree with Riley and Briars & Larkin and suggest that a child can apply procedures without having the necessary conceptual knowledge, and that conceptual understanding can emerge through the application of procedures. From a broader viewpoint, Davis (op.cit.) would emphasise the need for conceptual understanding. He suggests that the overreliance on procedural knowledge in mathematics has led to even highly educated students entering college with a flawed understanding of mathematics. He cites evidence from what has become known as "disaster studies". The term "disaster" is used by this researcher to reflect the fact that the supposedly "successful" students involved in the study have managed to progress with an essentially flawed understanding of rudimentary aspects of mathematics. These studies, carried out by Erlwanger and many others, show a number of misconceptions amongst these students. An example quoted is that of the student convinced that $\frac{2}{10} = 2.10$

This type of misunderstanding Davis suggests stems from a lack of proper conceptual understanding.

This is far from a complete debate of the topic relating to the relationship between procedural and conceptual knowledge but will suffice to show that it is an as yet unresolved area of research. It is, however, of great importance to the understanding of how children learn mathematics and in particular about how it should be taught. It is interesting to note that Briars & Larkin and Riley's models are based on information from children that are pre-school or in the early school years, whilst Davis was considering students with several years of formal education.

It is suggested by many, Davis (op.cit.) being one of them, that formal education is too concerned with procedural knowledge at the expense of conceptual knowledge. Certainly the disaster studies would suggest this. Evidence from Van Lehn's (1983) studies also suggest this. Many of the errors or bugs identified by Van Lehn can probably be traced to a flawed conceptual understanding. However, as pointed out by Silver (op.cit.) it is the relationship between conceptual and procedural understanding that is of importance, so what is needed is to get a balance between the two rather than considering one as more important than the other. It is easier to teach procedures for achieving answers and therefore this type of knowledge has perhaps been given undue stress in formal education. However, it is probably fair to say, certainly Baroody & Ginsburg (op.cit.) would consider it: that procedural knowledge may actually help conceptual understanding - practising certain routines may allow patterns to be discerned that help encourage conceptual understanding. Also by having "automized" access to number facts probably frees the brain to attend to other aspects of the problem. Hiebert (1990) also accepts the importance of automatizing certain routines, but he also emphasises the role of reflection: only through reflection will the patterns of mathematics, that are in evidence in the routines, be made available for further knowledge construction. However, seeing children struggling with the formal subtraction algorithm, who can in fact solve the problem by an informal method, and also children quite able to perform it but not having any understanding of when to apply it, must lead to a questioning of the undue emphasis on procedural understanding in formal education. This point was raised in the Cockcroft report (op.cit.) where the greater access to calculators and computers was considered a factor that should be taken into account when teaching certain formal algorithms. It singled out the long division algorithm as one procedure that could perhaps be left untaught. It also emphasised that these modern tools would put far greater demands on the ability to apply mathematical skills in a wide variety of situations. Hence it is essential to know which routine to choose but not necessarily essential to be able to carry it out.

The debate about the relationship is likely to continue. Determining whether a child has achieved conceptual understanding is not an easy matter. As an example, consider Piaget's conservation experiments which showed that the majority of children could not be understanding the concept of conservation until about the age of seven. However, a number of researchers, notably Margaret Donaldson (op.cit.)

showed that changes in the presentation of the task showed that the children had some understanding of conservation. Procedural understanding is easier to access - if a child can successfully carry out the same task on a number of occasions procedural understanding can probably be assumed. However, in the classroom there are a number of dangers when examining the different types of understanding. When a child fails to solve a word problem, the failure may have occurred for a number of reasons: not being able to read/decode the problem; lack of understanding of the problem statement - thus a failure in conceptual knowledge; or lack in procedural knowledge. If the child fails at the stage of decoding the problem s/he will not have a chance of showing whether s/he possesses the required conceptual skills, if the child has not understood the problem statement, s/he will be unable to show if s/he possesses the required procedural skills.

2.3 Arithmetic word problems.

Nesher (1976) suggests, on reviewing the literature, that there were two different approaches to looking at word problems: the translational and the structural. A third strand is suggested, which has been researched by Nesher and followed up by several other researchers: analysis of the underlying semantics of the problem. These three main approaches will be discussed separately in the order given above. Linked to the translational approach, which deals mainly with the language aspect of word problems is the readability of a problem. Thus this will also be considered within the translational section. An attempt to explain what happens if the problem solver does not attend to the underlying semantic structure but rather attends to surface structures has been made in verbal cue/keyword theory. This could perhaps be seen as a "mini" theory emanating from both the structural and semantic approach and will be discussed after the semantic approach.

2.3.1 The translational approach.

Much of the research in this area stems from Kane's (1967) research. His main contention was that mathematical word problems contained a mixture of "languages": ordinary English (OE), mathematical English (ME), and ordinary English words and phrases which have a specific meaning in mathematics that does not correspond to its ordinary English meaning. ME words and phrases are those only used in mathematics, examples are hypotenuse, coefficient and parallelogram. These words are rarely met with outside the classroom, often they are of Greek or Latin origin which may not be familiar to the child. The fact that there is little likelihood of the words being used outwith the classroom means there is little opportunity to practice the use of the word. As Section 2.2.2 suggests, children develop a more wide-ranging understanding of words in a gradual fashion and probably do so in a setting where feedback from the listener provides them with evidence of correct usage. A different problem arises when words have different meanings in mathematics from ordinary English. Often the mathematical meaning is more precise. "Difference" or "difference between" seems to be such a phrase. Few children would have any difficulty picking out differences between say a blue book and a red book that were otherwise similar. Often differences in this type of setting focus on a clearly visible physical aspect of the objects involved. In arithmetic "difference between" refers to the difference in quantity between two given numbers. This meaning seems obtuse to many children. They instead focus on facts such as "one is larger/smaller". Maybe this is a more perceptually based interpretation and therefore ties in better with the ordinary English usage. This difficulty was observed within this research project and is also well documented by other researchers (Bell et al, op.cit.). Apart from these different types of "language" within the mathematical problem it is also different from ordinary prose text in that it contains less redundancy. Redundancy in language refers to the words and phrases that serve to reinforce statements or comments made. A statement such as: "it is hot, don't touch it, it might burn you, it'll hurt if you touch it", contains several statements that all reinforce the idea that something hot will hurt. Mathematical problems have very low redundancy, and if the problem statement cannot be understood there are few extra clues to help the reader. This problem has perhaps been compounded by recent trends with individualised materials. If an individualised course is to be effective a certain reading skill is required by the user. The response to this demand has often led to a simplification of the material, so that short problem statements are used and the possibility of redundancy is reduced.

Another aspect of a word problem, related to the translation of it, is that symbols other than letters have to be read and understood. This matter is discussed fully by Shuard & Rothery (1984). They suggest that in the main symbols within mathematics are written according to convention. This means that the shape of configuration of the actual symbol is unlikely to suggest the meaning. The main symbols, within word problems in the primary school, that a child is likely to meet are integers: however, they may also meet with fractions. These, particularly vulgar fractions, can cause considerable reading difficulties. Prose text is read from left to right. The fraction ½ is not read as 1, 2, and it is pronounced as one half. Added to this difficulty, this fraction may be written in a number of different ways or expressed as a decimal fraction: ½, $\frac{1}{2}$, 0.5.

Related to this is the "writing" of the standard algorithms that children are taught. When the problem has been translated and a suitable operation chosen, the number aspect of the problem has to be transformed or translated into the correct algorithm if a written solution is required. The standard algorithm is written vertically not horizontally like ordinary prose. A number of conventions hold here: in subtraction the larger number is normally written at the top and the smaller number below and the answer is written below. In division the answer is normally written above. All these conventions are useful forms of shorthand when the exact procedure is understood and followed. It often leads to dramatic mistakes when it is not. This is well documented in

Van Lehn's research (1983), but can also be easily observed in many classrooms. Chapters 5 and 6 of this research contain evidence of this.

In order to measure the difficulty of a text readability formula have been developed. These types of measures typically consider sentence and word lengths, and familiarity of the words. Attempts were made to apply these type of measures to mathematical texts. Kane (1970) and Shuard & Rothery (op.cit.) consider this issue and suggest that these measurements are not suitable for mathematical texts for a number of reasons: lack of redundancy in mathematical texts which suggests that short sentences are not necessarily the easiest, length of word is not necessarily a good guide, and the inability of these measures to consider non-word symbols. Readability tests taking into account the specific nature of mathematical texts have been developed. However, the one developed by Kane is only suited to the American market and therefore could not be used in Britain without revision. Kane, Byrne & Hater (as quoted by Shuard and Rothery, op.cit.) suggest "they (readability measures) be used in conjunction with the judgements of teachers, curriculum workers and specialists in mathematical education". It may be fair to suggest that an experienced teacher may be able to judge the suitability of a text without the use of such a measure as they have a greater awareness of the language abilities of their particular group of children. These types of measures could therefore be considered to be of limited value.

2.3.2 Structural variables approach.

This area has been much researched by the Stanford group which was set up by Suppes in the late sixties. Jerman and Rees (1972) and Jerman (1973) were amongst those that continued this research. The word problem is analysed into discrete variables and the difficulty of the problem, it is suggested, is due to a small number of these variables. Altogether nineteen variables were identified as affecting the difficulty of the word problem. The following were amongst variables considered: type of operation required, absence/presence of verbal cues, verbal cue acting as distractors (i.e. the word suggests an arithmetic operation contrary to the one required, for example, the use of more when subtraction is required), memory, recall (of number facts), and length of problem statement. Regression analysis was used to narrow the range of the variables that most affected the difficulty. The Jerman and Rees study (op.cit.) found that four out of these nineteen variables accounted for 87% of the variance. These variables were: multiplication, division, length and verbal cue distractor.

The research method here was either in a CAI (computer assisted instruction) setting or using paper and pencil tests. In the CAI situation the students did not have to carry out calculation, only indicate the operation to be performed. In the paper and pencil tests the calculations were completed. The results only utilised responses as correct or incorrect, no attempt was made to look at different types of error. A number of variables that may affect problem difficulty were not considered. Two such that may be of importance is syntax and familiarity of words used. Linville (1976) suggested that both syntax and vocabulary level could have an effect on the difficulty of word problems.

The word problems used by Jerman (op.cit.) were controlled in length so that the number of words in each statement was a multiple of three. Adjectives and definite articles were inserted to modify the length of the problem statements. This gave three different sets of problems: one with the original problem length statement; one that was a third longer than the original; and one that was a third shorter than the original. All other variables were held constant. However, it may be that determining the length of the problem statement by the means of inserting extra words could have an effect on reading of the problem. It could create problems that seem more unnatural. It is interesting here to note the results from a study by Nesher (op.cit.). She looked at the effect of three structural variables: (i) Number of steps, the number of binary operations required to obtain a solution; (ii) Superfluous information, the absence/presence of superfluous numerical data in the word problem; and (iii) Verbal Cue, the absence/presence of word which may indicate choice of operation. A fourth variable: (iv) Question was also included which related to the type of story that was used as the setting for the problem. Variables (i), (ii) and (iv) were all found to have a significant influence on the children's ability to solve the problems, whilst the role of variable (iii) was not significant. An interesting result was that the responses to one of the questions deviated considerably from the rest. The researcher suggests that this may be due to the unrealistic manner in which the problem statement was worded. The problem statements were controlled for number of words and as in the Jerman study this may create a word problem that does not read well. The intention of a word problem is to allow application of mathematics in a natural setting. If the problem statement becomes very contrived this natural setting may not be achieved. It is interesting to note that Nesher has moved to a seemingly different perspective on word problems which builds on the structural approach but also looks at the semantic structure of the word problem. This approach is considered below.

2.3.3 The semantic approach to studying word problems.

This approach considers the underlying semantics of a word problem and discusses it in relation to children's interpretation of the problem. Nesher & Katriel (1977) consider addition and subtraction word problems in an article called "A semantic analysis of addition and subtraction word problems in arithmetic". They consider their approach different from previous research in that they: (i) emphasise "the semantic level of linguistic analysis"; and (ii) "that we characterize these problems in terms of their overall textual coherence, rather than focusing our attention on smaller linguistic units" (p. 252). The emphasis on "overall textual coherence" does suggest a move from the strictly structural which seems to consider problems broken down into subsections of the problem without the overall effect being considered. However, they still build on previous research by employing information processing type of explanations for their data. This same article goes on to propose that word problems are in some ways similar to riddles and that they conform to certain rules which are specific to word problems, in that they tend to employ certain phrases and a specific type of sentence construction. What is focused upon in this approach is the underlying semantic relationships of the problem. Many researchers (e.g. Carpenter, Moser & Bebout, op.cit., De Corte & Verschaffel, 1989, Riley, Greeno & Heller, op.cit.) have investigated these relationships and computer simulation models have been produced. It forms the basis for the part/whole theory discussed in Section 2.2.4 and this aspect of the approach will not be discussed further here.

De Corte & Verschaffel (1989) have suggested a model of competent word problem solving. It consists of five stages:

1. A complex goal orientated text-processing activity occurs which allows the problem solver to create an internal representation of the problem. This representation is based on problem sets and their relations.

2. On the basis of this representation the problem solver selects an appropriate strategy.

3. The selected strategy is executed.

4. The initial representation is reactivated to aid in answer formulation.
5. The answer is checked or verified to see that the answer ties in with the original problem.

Here it can be seen that in the initial stage, with the sets and subsets forming the basis for the problem representation, it is the semantics of the problem that is being attended to. This also ties in with the problem schemata of the part/whole theory, and, like Riley et al (op.cit.) De Corte & Verschaffel suggest that this stage is crucial to understanding problem solving. So, what is being proposed by this approach is that central to problem solving is the problem representation. Nesher, Greeno & Riley (1982) have developed a categorisation system of addition and subtraction arithmetic word problems that reflects the underlying semantic structure of the problem. They have identified four main categories: Change, equalizing, combine and compare. The first two involve actions that either increase or decrease a particular quantity; the last two are static in that they involve comparisons between two quantities. According to these researchers, and many following them (e.g. Carpenter, De Corte & Verschaffel, Riley et al) research has shown evidence of these semantic structures and computer simulation models (as described in Section 2.2.4) have been created to test the feasibility of these type of structures.

Verbal cue theory. It is interesting to note that both Nesher & Teubal (1975) and De Corte & Verschaffel (1987) use aspects of the structural approach to explain certain types of strategies used by children to solve word problems. The verbal cue or keyword strategy is used to explain how children can avoid creating a representation based on the semantic structure of the problem. Nesher & Teubal (op.cit.) suggests three levels of a verbal problem:

Level a - the verbal formulation

Level b - the underlying mathematical relation (this would presumably correspond roughly with De Corte & Verschaffel's Stage 1)

Level c - the symbolic mathematical expression (a correspondence to Stage 2 could be suggested).

Nesher & Teubal suggest that it may be possible to bypass level (b) if the verbal formulation [level (a)] contains a cue to possible action. Thus a word such as "more" may indicate addition. The problem solver would then move straight from the cue word in level (a) to a symbolic mathematical representation at level (c). At times this proves a successful strategy with correct solutions being achieved. However, when a word that may at times act as a cue for a particular operation occurs where that operation is not required, it may lead to an incorrect solution. An example, of this is a problem involving comparison using the phrase "how many more". The correct operation here is usually subtraction. If undue attention is given to the word "more" the child may choose addition.

Verbal cue theory fits into schema theory in that it could be explained as the application of an inappropriate schema.

Whether it is the best way of explaining the behaviour of the problem solver remains to be seen. This research project found no conclusive evidence for consistent use of verbal cue amongst the children studied (this is discussed in Chapter 6).

The semantic approach drawing on information processing models has been criticised by Lean, Clements & Del Campo (1990). They suggest that psycholinguistic models based on those of Clark, Donaldson & Balfour and others provide explanations that are equally adequate. They suggest that the children's understanding of what they term "polarised comparative terms" - terms such as "more-less", "big-small" and "same-different" - is not well established in the early school years. This lack of precise understanding is preventing the children from coping with word problems. They suggest that many textbook writers are unaware of this lack of linguistic skills in many young children and that word problems frequently use linguistically highly complex constructions. The numeracy demands of these problems are often simple and as word problems seem mainly to be seen as a vehicle for developing arithmetic process skills the language demands tend to be ignored. Lean et al suggest three different classes of strategies are employed when solving word problems using the words "more" or "less":

Type 1	(for more)	involves finding the larger number and giving this as the answer
	(for less)	the smaller number is given
Type 2	(for more)	an additive process is applied

(for less)	a subtractive process is applied
Type 3 (for more and less)	a mental representation is created and an arithmetic operation based on this representation is performed.

The Type 2 strategy contains certain similarities to the verbal cue theory discussed above. Lean et al, however, maintain that children's understanding of the phrases more and less can be so precarious that many of them may interpret more as less and vice versa. This is not an explicit assumption of verbal cue theory, and it certainly makes it considerably more difficult to work out if a child has in fact responded to the question in a verbal cue manner. In support of verbal cue can be cited that Lean et al found that Type 2 errors were generally in the direction of children adding when confronted with "more" and subtracting when encountering "less". Perhaps it should be mentioned here that Nesher & Katriel (op.cot.) carried out their research using children in their teens whilst Lean et al are concerned with children from the age of five. Whilst they report that Type 1 kind of errors were most frequent among the younger age groups they were found amongst children up to the age of eleven. Lean et al do not mention whether any children in the older age groups still believed that more meant less.

Another interesting point, which again supports general language schema theory, was that in individual interviews it became apparent that the children frequently imposed a meaning on the question that was different from the actual question. An example quoted is the following question:

"Nick has 2 bottles. Jane has 7 bottles. How many bottles less than Jane does Nick have?" An able twelve-year-old responded with the answer "Nick". Questioning revealed that he had interpreted the question as "Who has less bottles, Jane or Nick". A similar "reinterpretation" of the question has been noted in this research project where many children inserted an "and" between two separate questions and thereby creating a different problem, or where certain parts of information were left out and thereby simplifying the problem (see Chapter 6, Section 6.2.2).

Lean et al thus suggest that the specific theory emphasising semantics may not be the best theory to explain children's mathematical understanding. However, general schema theory is still employed as an overarching general theory by both the approaches. It remains for future research to consider which of the two - the semantic model, based on part/whole theory, or the psycholinguistic model - provides the better explanation of children's arithmetic word problem solving behaviour. That further research is important is evident by the fact that the two different approaches advocate different application of their research to the classroom situation. De Corte & Verschaffel (1989) feel that a far greater variety - in terms of underlying semantics - of word problems should be given to young children. Lean et al urge caution and suggest that it is essential that the language involved should be kept at relatively simple level, and that teachers need to ensure that the language used is understood in the intended manner.

2.4 Features of textbook organisation that may affect word problem solution.

The previous section looked at research relating to children's ability to solve word problems. This type of research has mainly used researcher created problems to control carefully the variables under study. Page presentation has thus not been considered. Shuard & Rothery (op. cit.) have included a discussion on these features but little research evidence is available. Common sense would suggest that page layout should be clear and easy to follow and any colour used should not obscure the text. Probably the most important feature external to the problem statement is graphics. Word problems often make use of graphs, maps and pictorial illustrations. There seems to be no research in this area relating to the age group used in this project. However, in the present study, both layout and essential and non-essential pictorial illustration had an effect on the children's ability to solve some of the problems. These effects are discussed in chapters 5 and 6 and will therefore not be repeated here. It does seem, though, that future research on word problems should consider the word problem in its totality - that is how it appears on the textbook page - rather than concentrate on isolated word problems.

2.5 The effects of schooling and the child's background.

Home background and its effect on educational success is an area that has been greatly researched. It has not been the focus of this study but if the child is considered an active participant in the learning process, as suggested by schema theory, then it cannot be ignored. A brief discussion will thus be included here and it will focus on language factors. The reason for this focus is that whilst mathematical understanding may not be stored in language form, language forms the main medium for transferring knowledge from one individual to another, and does depend on the child's background. Thus in the early years of learning the formal system of mathematics it is likely to be of great importance.

The debate in this area has centred around the language deficit/difference concepts. Bernstein (1979) proposed that children from a lower social background had access to more limited language structures than had middle class children. Thus he argued they found it difficult to make sense of the mainly middle class, elaborated code of the classroom and educational failure was more likely. Labov (1979) amongst others argued against this standpoint. His research suggested that lower class children did not have more limited structures, rather these structures were different but equally rich compared to the middle class structures. The difficulty that many lower class children were experiencing were thus not due to lack of language. Mismatch between the child's language and that of the school was considered a possible cause of difficulty for these children. What is of importance here is that language used in mathematics is very precise and the teacher needs to be aware of the mismatches that might occur if the child is not interpreting a word in the manner intended and expected. In this study no attempt has been made to study these variables. However, they are nonetheless important and should be borne in mind by any researcher studying children's classroom learning.

The teacher and institutional variables have as home background been peripheral to this study. Again, like home background they are not unimportant, only a research project of this nature needs to focus on a limited number of variables. As the teacher mediates in the learning of mathematics a brief discussion of these variables is in order - again there is much research in this area and only a few aspects that are considered of particular importance will be discussed. This project involved five different schools. Differences were found between the different classes, suggesting possible teacher/school effects. These are discussed in Chapters 5 and 6 and will not be further discussed here. What could usefully be considered is what might be termed a paradigm shift within teaching methods and how this effects the teacher in the classroom.

The earlier paradigm was based on the behaviourist paradigm as discussed in Section 2.2.1. Children learnt by association; rote learning and drill would improve their ability to do mathematics - this is also sometimes referred to as the traditional approach. The new paradigm based on schema theory suggests that the child is an active participant in the learning process. What the child brings to the learning situation is an integral part of how it is learned. This is discussed in Section 2.2.2. This change means that the teacher is faced with the need to provide suitable learning experiences for a large number of children who may have quite different backgrounds. What can be seen in these shifts in methods is that classroom organisation needs to change to accommodate them. Whole class teaching is not considered acceptable as the main tool of instruction; the focus has moved to smaller groups and individuals. The debate as to the efficacy of traditional versus modern methods is still raging. It was discussed in chapter 1 and will not be continued further here. It can be seen though that the teacher is faced with a very difficult job - that of creating optimum learning conditions for a large number of children many of them at different stages of learning mathematics. There has been a proliferation of schemes, textbooks and sets of materials that are intended to increase the ability of the teacher to provide individualised learning. If one accepts the tenets of schema theory that learning, particularly in the early stages, involves a "negotiation of meanings" which is largely dependent on language it becomes difficult to accept these schemes as providing an effective solution to the problem. It is not easy to "negotiate" meaning with a written text if that text is written in language that does not convey much meaning to the reader.

The concern expressed by many researchers that schools are insensitive to individual children's needs [see for example Baroody & Ginsburg (op. cit.) and Davis (op. cit.)] is not unfounded. Undoubtedly there is room for improvement in some classrooms. It may be that under present institutional constraints, both in terms of school organisation and further training opportunities, only limited improvements can be made. Desforges (op.cit.) also feels that social scientists should "direct their prescriptions for learning not at teachers but at those who provide for and plan the conditions of classroom life" (p. 292) Perhaps the one class - one teacher concept is not useful given modern teaching methods. Also, given that children develop at different rates maybe the age related class structure is not effective Dr. Mary Simpson of Northern College of Education suggests this (Henderson, D. 12.4.91) : "perhaps we need to get rid of the concept of the class and break with the idea that by treating people alike, because they are the same age, regardless of manifest differences, we have a good model for equity and equality: we haven't. And this model persists nowhere but in schools" (p. 5). Mary Simpson implies that an absurd system may be the cause of lack of success - the results of this study suggest that this may be too simplistic.

This brief discussion has considered variables that are of importance in children's learning of mathematics. Future research will have to consider their influences in greater detail.

2.6 Conclusion.

This chapter has traced changes and developments within the psychology of learning that show how the perception of the learner has changed. The child is now seen as an active constructor or his/her knowledge of mathematics, in fact of any knowledge. Schema theory is now accepted by many as an acceptable model of how an individual, child or adult, organises the knowledge s/he has of the total environment. Within mathematics education it provides a valid framework and has led to research that looks at the development of more specific schemata for mathematical understanding. This research project finds it a useful model for explaining children's understanding (or lack of it) of word problems.

In the future, schema theory in relation to mathematics learning may need to consider the different types of schemata that would be needed to account for the different types of knowledge: conceptual and procedural. A beginning has been made in part/whole theory with the discussion of the problem and the action schema. According to Baroody & Ginsburg (op.cit.) procedural knowledge can precede conceptual knowledge in the developing understanding of mathematics. However, it may be that as higher levels of mathematical understanding are encountered it becomes essential for procedural schemata to become linked to conceptual schemata in a subordinate role.

Looking at the research relating to the actual word problem, it is clear that by breaking the problem into smaller parts, the approach of the structuralists, and by looking at a limited area of development much has been learned and will continue to be learned. However, the larger aspects must also be considered and it is important that research in particular areas does not become isolated from the rest. Arithmetic word problems occur in a setting - that of the textbook page and that of the classroom. These "external" aspects of the problem have not been considered by many researchers. This particular piece of research aims to look at the word problem in an "ecologically valid" setting - that of the classroom using the textbook currently used by the majority of Scottish classrooms. As previous researchers have found, limitations always have to be imposed, and that has been the case in this study. Only a small number of actual word problems have been included in the study from the vast range that the children meet during the school year. It is hoped, however, that by studying these few problems intensively, using a variety of methods, and in the school setting something can be learned that will further our understanding of how children approach the task of solving word problems. It is to the study of the children's interaction with these word problem that the following chapters are devoted.

CHAPTER 3 – RESEARCH METHODS AND APPROACHES

3.1 Introduction.

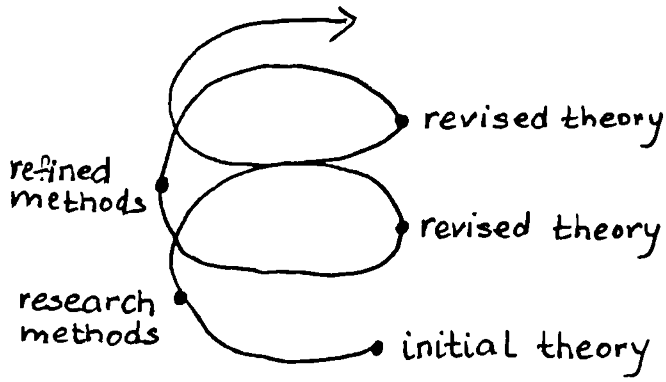
The last chapter examined the theoretical background to this research project. This chapter will look at the research methods available within educational research and explain the choice of methods used within this project. The chapter consists of four main sections:

(i) an overview of research methods in the social sciences
(ii) data collection methods chosen for this project
(iii) data analysis techniques used in this study
(iv) the exploratory study testing the chosen methods

3.2 An overview of research methods and approaches.

To set in context the methods used in this project, and to explain their choice, this section discusses some of the research methods available to educational researchers. The basic intention of any research is to examine methodically the phenomenon being studied - in the social sciences this "phenomenon" to be studied is either the individual or groups of individuals. However, the methods that have been used within educational research have varied. The previous chapter, in Section 2.2, looked at two theories that set out to explain human behaviour and learning: the behaviourist theory and schema theory. These theories have influenced educational research and have affected the different types of methods being developed in this field of study. In Section 2.5 it was suggested that a shift in paradigms from the behaviourist/"formal" to the "informal", based on schema theory, had affected classroom organisation. This shift in paradigm can usefully be employed here when looking at research methods. The methods employed are chosen in order that they can usefully explore some aspect of a particular theory. However, changes in method are not solely due to paradigm shifts but also to increasing technological advances, making data gathering easier and widening its scope. These tools will be discussed in relation to the methods that employ them.

Perhaps the relationship between theory and method can usefully be summed up using a helix to illuminate the continuing interrelationship:



At the early stage of theory formation ideas and questions require exploration - this leads to suitable selection of methods - theory is revised - methods may be refined/changed to help provide answers to those requiring attention as indicated by the revised theory.

This type of development was shown by Piaget's work. He pioneered the use of the clinical interview with young children. This yielded rich data which provided explanations for a developmental theory. As this theory developed, other researchers questioned some aspects of it and tried out revised methods.

3.2.1 Behaviourist - quantitative - research methods.

As discussed in Chapter 2, the behaviourists limited their study to observable, external aspects of behaviour. The end product, the response, was of main importance to these researchers. Hence methods that provided responses for analysis were used. The main one of these was paper and pencil tests. They were easy to administer to large numbers and easy to score. They were considered reasonably easy to write. In order to retain objectivity, for example when looking at memory, nonsense syllables were used. It was assumed that these would be equally unfamiliar to all subjects. Within mathematics, a correct response was considered to show understanding of the concepts involved. Added to these earlier techniques were statistical measurements that allowed for comparisons between different groups - groups such as different classes, different types of pupils or types of problems. A statistical relationship between two variables is taken as indicating a relationship between these two variables. Much of our educational examination system still use similar techniques, perhaps particularly so within mathematics education where it is relatively easy to set questions that have only one correct answer. These type of methods have become known as the "quantitative" approach for the reason that they attempted to quantify behaviour so that it could be measured easily. From this data general trends could be suggested of how certain groups would behave, be it within the learning of mathematics or in any other situation.

For the behaviourist researcher technological advances have made more precise measurement possible: timing devices allow for accurate timing of responses; devices that trace eye movements can focus on a subjects eye movement through a text; and tachistoscopes allow for time-controlled presentation of written material and precise timing of response times. The use of the first two of these devices within mathematics education research will be discussed below in section 3.2.2.

3.2.2 Qualitative research methods.

The quantitative methods discussed above provided considerable amounts of data on large numbers of people. Coupled with statistical techniques, general trends could be discerned. However, they do not necessarily provide much information on the particular behaviour of an individual. For example, an incorrect response on a mathematics test may be due to a number of factors: incorrect reading of the question, not understanding the question, missing out the question by mistake, not being able to carry out the computational aspect of the problem, or not recording the answer correctly. All these errors might be labelled the same - "incorrect" - by the quantitative approach. This lack of differentiation was one of the reasons that some researchers turned to different methods. Davis (op.cit.) refers to the multiple choice achievement tests of the quantitative researchers as "akin to peeping through a keyhole to find out what is going on in a room". What is needed, it has been argued, is a change from analysing the end product to investigating the processes that people employ to reach a solution. By looking at the processes it may be possible to see where misconceptions arise and thus create a chance to correct these misconceptions.

In order to investigate processes in mathematical understanding protocol methods have been developed by educational researchers. There are two different types: talking aloud and clinical interviews. Similar to the clinical interview is the task-based interview and the two will be discussed in conjunction with each other.

The "talking aloud" method works on the principle that the subject is asked to talk aloud as s/he attempts to solve a problem. The assumption is that in doing so thought processes will be revealed for the researcher to record and analyse further. Newell (1977) describes some of Newell and Simon's work which pioneered this method. This method has been used mainly with adults as children, particularly younger ones, often find it difficult to articulate their thoughts. The main role of the experimenter is to explain the task to the subject. The subject then "verbalises" his/her solution and this verbalisation is recorded.

The "clinical interview" stands in contrast to the "talking aloud" method in that it questions the subject, as required, whilst the task is being solved. As mentioned, this method is more commonly used with younger children. Questioning is usually related to the reason for particular steps being taken in the solution process. It is a method much used by Ginsburg and his associates (Ginsburg & Allardice, 1983), and has been an effective way of gaining further understanding about individual learning difficulties in mathematics.

Similar to the "clinical interview" is the "task-based interview". The role of the experimenter is the same in that s/he attempts to increase understanding of mathematical development. The method is described by Davis (op.cit.). The main difference is that the clinical interview tends to be used in a setting where the interviewer/experimenter is trying to understand individual difficulties. Thus the interviewer would focus on aspects of mathematics that the particular individual involved finds difficult. The task-based interview is used to find out how individuals are likely to carry out a particular task that reflects some aspect of mathematical understanding, regardless of whether this task is considered difficult or easy. In this situation the interviewer asks the child to solve a pre-determined task. Questions may be asked to illuminate the solution process as required. Many of the researchers looking at the semantics underlying word problems (see Chapter 2, Section

2.3.3) have used this type of method to investigate children's understanding of simple arithmetic word problems.

The qualitative methods described above have provided researchers with a wealth of data on problem solving behaviour in several areas of mathematics. They are criticised by some for not being objective records of an individual's behaviour. The interviewer may choose to follow a particular path in the interview or to change as the situation demands - this is a subjective judgement. It is difficult to recreate an interview situation, so it becomes difficult to carry out replication experiments to test the findings. However, these methods do provide interesting data and provided all aspects such as intended interview schedules and any deviations are recorded it is possible to check the data. Recording the data on tape also provides for a possibility of analysing the data after the interview and for other researchers to reanalyse.

The limitations of both quantitative and qualitative methods have led a number of researchers to use a combination of methods. De Corte and Verschaffel (1989) use what they consider a "broad spectrum" approach. The initial method used by these researchers whilst investigating children's understanding of arithmetic word problems was the individual or task-based interview. Added to this method has been paper and pencil tests, observation of behaviour during solution, eye-movement registration and teaching experiments. The paper and pencil tests administered to larger groups have allowed problem difficulty to be determined. Eye movement registration has suggested differences in fixation time between successful and unsuccessful problem solvers.

Technological developments in general have allowed for increased sophistication in educational research. Although tape recorders have been available for some time the availability of relatively cheap, unobtrusive and easy to use recorders have made it far easier to carry out research in the natural setting. Analyses of the material can then be carried out at a later time. This leaves the researcher free to observe more closely the actual behaviour of the child. Video recorders are also used but not the same extent due to the cost and more cumbersome equipment. Eye movement registration is another device mentioned. Again this requires expensive equipment and cannot easily be used outside the laboratory. This method has shown interesting differences between successful and unsuccessful problem solvers (see De Corte & Verschaffel, 1989). The main difference between these two types of learners is that the successful problem solver spends more time focusing on the important parts of the problem during the decoding stage.

So, educational research since the 50's has changed from using mainly large scale paper and pencil type tests to more probing interviewing type techniques. The old methods are still in use but are complemented by methods that are more effective at identifying individual differences. Technological advances has proved an aid to data collection and analysis.

3.3 Data gathering methods chosen for this research project.

The aim of this research was to investigate the types of difficulties that children experience when trying to solve arithmetic word problems in their standard textbook - SPMG Stage II. There are three identifiable elements to the project: an exploratory study to pilot methods and reconnoitre the area; Phase I to identify potential sources of difficulty; and Phase II to test the effects of linguistic and computational changes of the word problem. Phase I of the project is reported in Chapter 5 and Phase II is contained in Chapter 6. Two methods were chosen: the task-based interview, and paper and pencil tests - standardised and researcher created. The task-based interview was used during Phase I of the main study to ascertain problem difficulty. Paper and pencil tests were used to test the effect of the structural changes to the word problem during Phase II.

3.3.1 The task-based interview. This method was chosen to identify the types of difficulties that the children experienced when solving the selected problems. The term task-based rather than clinical interview is used here to reflect the fact that all the children were presented with the same tasks irrespective of the types of difficulties they showed. The aim was to investigate difficulties inherent in the chosen problems rather than remedying individual pupils difficulties. Two aspects need to be considered here:

(i) the tasks chosen

(ii) the interview format

(i) The tasks chosen. It was intended in this project to look at the type of arithmetic word problems that the child is likely to meet during his/her ordinary school work. The SPMG scheme is the main mathematics scheme in the majority of Scottish schools and the schools participating use it as its main, though not necessarily only, scheme. The problems used were chosen at random from the early part of the book. This was to ensure that the children involved had been taught the methods required for solving the problems. Most of the children had already encountered the problems involved during the ordinary work of the class. The final selection of word problems is discussed further in Chapter 4, Section 4.2.2 (a).

(ii) The interview format. It was explained to the child that the researcher was interested in finding out how children do word problems. To find out a bit more about this the child was going to be asked to solve some word problems from Stage II textbook. This explanation was followed by a simple request to do a particular problem. A framework of prompts was created to be used if the child was unable to progress towards a solution. Newman's classification of errors as described by Watson (1980) was used as a basis for this framework. This would ensure that all children were exposed to a similar type of treatment. To make it exactly the same is not possible in this type of situation as the children were not all experiencing the same type of difficulties. Newman's classification consists of eight steps:

a) reading the problem
b) comprehending the problem
c) transforming the problem
d) the mathematical calculation
e) encoding the answer
f) motivation
g) carelessness
h) question form

Out of these eight steps the first five were used. The final three, it has been suggested by Watson, do not have the same impact on younger children and are not of such importance in an interviewing situation.
The interview format consisted of asking the child to do the selected word problem. If the child was stuck, s/he was asked to read it. If a difficulty was evident here the researcher helped the child read the problem. If this was not sufficient comprehension was checked, followed by help with transforming the problem, should it be necessary. It was then possible to record if the mathematical process skills were present or absent. To encourage suitable encoding a simple question "is that all?" was used. This interview format provided a loose framework for the task-based interview. The procedure is further discussed in Chapter 4 Section 4.2.6 and 4.4 (1).

3.3.2 Tests used within this project. Two types of tests were used: standardised, commercially produced tests; and researcher-created tests covering aspects of mathematics. Tests were used for three reasons: (i) to compare children's understanding of mathematics and language; (ii) to select a sample population in the main study; and (iii) to investigate the effect of structural problem changes on problem difficulty.

(i) Tests used in comparing children's understanding of mathematics and their linguistic ability. Initially this project set out to investigate links between linguistic ability and mathematical difficulties. In order to look at this possible relationship the pilot study tested the children's computational ability and reading comprehension.

The computational ability test was created using the 1984 Assessment of Achievement material for guidance. This material has been developed for the Scottish school population and was therefore deemed a suitable basis for an arithmetic test. As a distinction between numerical ability and linguistic ability was sought it was felt necessary to limit this test mainly to computation with only a few word problems included. This would limit the language demands and allow the child who may be disadvantaged linguistically to show any numerical ability.

To test language ability the Edinburgh Reading Test

(1977) was used. It is a standardised, commercially produced test. It has been written for and tested on a Scottish population. It includes measurement of vocabulary, sentence comprehension and use of context and is thus more wide-ranging than many other commercially produced tests.

The intention, initially had been to investigate the possibility of pupils who may be disadvantaged in mathematics due to lack of linguistic skills. If the scores attained in the two tests were divided into two - high and low - based on the scores in the two tests the emergence of four groups were envisaged, showing the language test scores first: high/high; high/low; low/high; and low/low. Of particular interest to such an investigation would be the pupil scoring low in language and high in computation. The exploratory study described in chapter 4, was used to test the feasibility of the task-based interview in a classroom situation but also to look for evidence of these types of groupings. It was found as the project progressed that arithmetic word problems consist of a complex interrelated set of skills, and that it is difficult to disentangle language from mathematical skills in a clear cut fashion that would be useful from a classroom point of view. There were undoubtedly a small number of children hampered in mathematics due to their inability to use language effectively. However, many more were affected by a lack of understanding of mathematical conventions, be they linguistic or based on number symbols, and it therefore seemed more pertinent to explore some aspect of this relationship.

(ii) Tests used in selection of a sample population for the main study - Phase I and II. The exploratory study investigated the type of difficulties children experienced when solving word problems. The task-based interview provided the tool for accessing this information. The findings from the exploratory study suggested that it would be useful to investigate these difficulties in a larger number of schools targeting the "average" pupil. To select this group a standardised mathematics test was chosen. Unfortunately no mathematics test standardised on the Scottish school population is available. The choice and suitability of the chosen test is further discussed in Chapter 5, Section 5.2.2 (c).

(iii) Tests used in investigation of the effects of word problem structure on word problem difficulty. For this purpose researcher created tests were used. The development of these tests was based on the type of difficulties found during the task-based interviews during Phase I. The selection of problems and creation of further problems is described fully in Chapter 6, Section 6.2.2.

In conclusion, then, the research methods chosen for this project are similar to those of De Corte and Verschaffel

(1989) in that they employ a mixture of methods. It has, however, been necessary to leave out those aspect requiring complex equipment such as eye movement testing. A further limitation is that only one person was involved. Visiting five different school and interviewing forty children is time consuming - one of the main drawbacks of the interview method. However, this choice of methods have provided interesting and useful insights into children's problem solving behaviour in an ecologically valid setting.

3.4 Data analysis techniques used in this study.

As suggested by the overview of research methods quantitative methods have tended to use tests that are easy to score. Frequently a response is simply scored as correct or incorrect. This provides an indication of problem difficulty and has been used here. However, the purpose of the task-based interview was to provide deeper understanding of the difficulties experienced. A more fine-grained analysis of the incorrect responses are therefore necessary. Watson's (op.cit.) adaptation of Newman's hierarchy of errors was used for the interview prompts. This error analysis was also applied to the data collected in the task-based interviews. The initial adaptation and use of the error analysis is described in Chapter 4, Section 4.3. Suggested changes are discussed in 4.4 (3) and in Chapter 5, Section 5.2.7.

3.5 Exploratory trial of the chosen methods.

The methods chosen for the project have been discussed in the preceding sections. It was decided that it would be useful to test out the feasibility of using the task-based interviews in a classroom setting. The selection of a sample population could also be explored. A study by Nicholson (1977) in the secondary school suggested difficulties with mathematical terms not only amongst poor learners. He suggested that as many as 50% of the middle ability range have problems understanding mathematical terms in common use. It was therefore felt that the middle ability range might provide an interesting area of study. However, it was felt necessary, initially, to look at the whole ability range within a class to provide an indication of the variation that may exist within a class. This formed the exploratory study. It also provided a vehicle for selecting a number of word problems suitable for further study.

3.6 Conclusion.

This chapter has provided a brief overview of the research methods available to educational researchers. It was used to explain and set into context the methods chosen for this particular research. The analysis of data and testing of the methods through the exploratory study have been discussed and specific links with the chapters to come have been indicated.

CHAPTER 4 – EXPLORATORY STUDY

4.1 Introduction.

This is an account of the project's exploratory study. The main aim was to explore the feasibility of using the task-based interview, in a school setting, as a tool for exploring children's understanding of word problems in a textbook. A subsidiary aim was to examine the possibility that some children may be experiencing difficulties in mathematics due to lack of language skills. This chapter will follow, broadly, the format of an experimental report with the following sections:

(i) method, including details of design, materials used, school background, subjects, apparatus and procedure;
(ii) error analysis technique;
(iii) results and discussion; and
(iv) conclusion, including the consequent rationale that underpinned the following main stages of the project.

4.2 Method.

4.2.1 Design. As discussed, the intention of this exploratory study was to examine how P4 children cope with certain word problems in SPMG textbook Stage 2, and then to examine the types of difficulties experienced. Task based, tape-recorded interviews, with each child being interviewed individually, formed the core of the study. The tasks consisted of sixteen SPMG word problems and the child was observed and interviewed whilst carrying out the task and on completion of each task. The interviews were intentionally open-ended and the children were helped over any difficulties they experienced to reach a solution to the problem. The actual problems used for the analysis are listed in Appendix A. To aid in the selection of a suitable sample, a mathematics test was used in conjunction with the teacher's ability groupings. In order to look at any possible relationships between ability to solve word problems and language skills a language comprehension test was administered to the whole class.

4.2.2 Materials. (a) SPMG textbook, paper and pencil
(b) mathematics test
(c) language test.

(a) Sixteen word problems from SPMG textbook Stage 2 were used in the analysis (see App. A). Final selection of word problems was deferred until the initial interviews had indicated the type of problems that would provide useful data. This was to allow for exploration of different types of problems. It was not known how long it would take for each child to complete the problems and a great deal of variation was expected here. In class the children tended to work with other pupils and those not quite competent were being helped by those able to do the problems. In the interview setting each child was on his/her own. Originally it had been intended to work backwards roughly from the stage the children had reached in the book. However, levels of competence did not match the stage reached, and only the most skilled children were able to cope with the problems thus selected. It was felt that if the problems used were beyond most of the children they would be put off trying to solve them. The strategy was therefore changed and problems were selected that started at page 1 of the book and worked forward to include a number of problems covering the four arithmetic operations. No attempt was made at this stage to select an equal number of problems from each arithmetic operation; rather diversity was sought in operation, graphical presentation and layout of the problem. This was done in order to see if there were any particular type of word problems that created specific difficulties. These problems were then used in further studies and were altered in a number of ways to examine the effect of these alterations on the pupils' ability to achieve a solution. This was an exploratory study, and accordingly there were no specific hypothesised expectations about the difficulties of the problems. It was, however, expected that addition problems would prove the easiest, and that subtraction and multiplication problems would be more difficult, and that division would be the most difficult.

(b) The mathematical test used was created from some parts of the Assessment of Achievement Programme (AAP) 1983 for Primary 4. It had been intended to use a standardised, computational skills test originally. However, after looking at the tests available and consulting Ridgway's "A review of mathematical tests" (op.cit.) it was found that there was a dearth of suitable tests with standardisation for a Scottish population. The choice was then made to create a test based on information available from the AAP 1983 Primary 4 test. This test is included in Appendix B.

(c) The language test used was the Edinburgh Reading Test, Stage 1 (1977).

4.2.3 School/classroom background. The school where this study was carried out was situated in a small Scottish city. The area was predominantly middle-class. The school had just over 200 pupils with no composite classes. The primary 4 class involved had thirty-two pupils and a male teacher. The main means of instruction in mathematics was the workbooks and the textbook from SPMG Stage 2. There was some expository instruction in specific topics such as symmetry or division. The main part of the time spent on mathematics consisted of individuals working through the textbook or workbook material. Talk was allowed, so many of the pupils worked together with a neighbour. Any pupil requiring help would call the teacher's attention and get help individually. This was naturally rather time consuming for the teacher; he commented that he felt he was not able to spend as much time with individual pupils as he would like. Occasionally a textbook from another scheme was used, mainly for the top group. The teacher felt this book provided more "mathematics" as it contained complete pages of exercises of the same nature, rather than the mixture that is common in

SPMG. Concrete materials in various forms were available on a separate table for the pupils' use. The atmosphere in the classroom was one of informality. For the children who found concentration difficult it may have been a little bit too noisy, with much to distract them from their work.

4.2.4 Subjects. The size of the sample was governed by the time available. The decision was to see as many subjects as possible within the four week period allowed for the study. The choice of subjects for inclusion in the study was affected by two factors: the teacher's ability groupings in the class (which consisted of three groups); and the results of the maths test. A small number of successful pupils were to be included among the population. These were selected by including those pupils from the top ability group with the highest scores on the maths test. There were a number of discrepancies between the teacher's groupings and the results of the maths test, and out of interest a number of low test scorers from the teacher's top ability group were included. Altogether five pupils from the top ability group took part initially in the study; however, one of these pupils moved away during the period of the study and was replaced by another pupil from the same teaching group, so only four complete transcripts are available from this group. The second group formed the slightly larger group and consisted of pupils of supposedly average ability. It was also considered to be the group of most interest for this project, so a larger number of pupils were included from this group. Its size not only reflected research interests but also the fact that the teacher's middle ability group was the largest group in the class. They were drawn at random from the teacher's middle group and included two pupils that had scored well above average in the maths test. Altogether eight pupils were included from this group. The lowest ability teaching group consisted of only six pupils and from this group two pupils were selected at random for this study, and one of these pupils did not complete all the word problems due to severe learning problems.

The ages of the children involved ranged from 8 years and 1 month to 10 years. The ten-year-old was the child with severe learning difficulties; the top age for the rest of the group was nine which gives an age span that is normal for any Scottish Primary 4 class.

4.2.5 Apparatus. A Sony TC-D3 stereo cassette recorder with a microphone was used.

4.2.6 Procedure.

(i) Task-based interviews. Each child was interviewed individually in the library, adjacent to the classroom. The format of the individual interviews has been outlined in detail in Chapter 3, Section 3.3.1. The library also formed the secretary's office and was occasionally used by other children carrying out project work. However, this did not seem to interfere with the childrens' concentration. They were used to working in a classroom where there was always plenty of activity. The tape recorder was used to record every interview. An effort was made to keep the interviews informal so that the child was relaxed. On the whole this seemed to work well. Each child was quite happy to carry out the tasks as requested. The interviews lasted for as long as it took the child to complete the tasks and were all carried out during the morning session. In a number of cases the interviews continued after the playtime break to complete the tasks. As mentioned earlier in this chapter, Section 4.2.2, the strategy for choosing word problems was changed as it became evident that the average child was not able to cope well with the word problems initially chosen. This led to a number of pupils not having carried out all the sixteen word problems. To remedy this, a second interview was conducted with those children. Only those problems not already attempted, out of the final sixteen chosen, were used on this occasion (see also Section 4.2.2). The children were all told that the researcher was interested in finding out more about how they did their maths and that they would be asked to solve a number of word problems to show how they did them. If they were unable to find a solution they would get help. They were also told that the interview would be recorded but that only the researcher would listen to it afterwards. The tape recorder was then started and the child was asked to start on the first problem.

The need for flexibility in the interview schedule was emphasised by different responses to the same interviewer probing question. Problem 3 read:

One morning 37 boys and 46 girls go to the library. That afternoon 39 boys and 59 girls.

How many (a) boys, (b) girls go to the library that day?

The interviewer probe "how did you know you had to add them up?" to examine why the child had chosen that particular action produced two different responses. One child responded with: "I did a 9 and a 7 and that's 16, carried the 1, and then 3 and 3 makes 6 and you add the 1, it makes 7". A different child responded with: "cause it says how many boys, it says on one morning 37 boys and then that afternoon 39 boys .. so you add the two together". Here the prompt has extracted from one child purely *what* she did in order to get the answer, from the other child *why* she selected a particular operation. Thus the same question does not always elicit the same type of answer and it becomes essential for the interviewer to progress differently with these two children.

(ii) Mathematics and Language Tests. The mathematics test was used to select a sample in conjunction with the teacher's grouping, and the language test was carried out in order to examine any possible links between difficulties in mathematics and language. The whole class was involved in doing both the tests. The testing was done in the classroom during the forenoon.

4.3 Error Analysis Technique.

Newman's error analysis. The technique used here was based on Newman's error hierarchy as described by Watson (op.cit.). The Newman hierarchy contains eight main categories with subdivisions within some of the categories. An outline of these categories is shown below:

Main category	Sub category
1. Reading: can the pupil read the question?	(i) word recognition (ii) symbol recognition
2. Comprehension: can the pupil understand the question?	(i) general understanding (ii) understanding specific terms
3. Transformation: can the pupil select the mathematical processes which are required to obtain a solution?	
4. Process skills: can the pupil perform the mathematical operation necessary for the task?	(i) random response (ii) wrong operation (iii) faulty algorithm (iv) faulty computation (v) no response

5. Encoding: can the pupil write the answer in an acceptable form?

6. Motivation: the pupil could have solved the problem correctly had s/he tried.

7. Carelessness: the pupil could do all the steps but made a careless error which is unlikely to be repeated.

8. Question form: the pupil makes an error because of the way the problem has been presented. (A question may be ambiguous, for example).

This error analysis was used by Newman and Clements on children in grades five to seven. Watson (op.cit.) carried out a study to see if this type of analysis would be useful with younger children. In his study the ages of the children ranged between 6½ and 7½. He found the error analysis feasible even for this age group and emphasised the possibility of its role as a diagnostic tool for the teacher. Categories 6 to 8 were not used by Watson as he felt there was no need for them with this age group. There are certain problems with these last three categories: motivation could be difficult to assess and, as Watson suggests, on the whole younger children are keen to work and this category becomes more relevant with older children. Carelessness is also problematic when it comes to assessment, and it would be essential to know a child's past performance well in order to use this category. Newman and Clement classified as careless a child who made a mistake on their written test but did not repeat it in the individual interview. However, the difference in conditions may have accounted for the child performing better in the interview situation and s/he may well make similar "careless" mistakes again in a written test. It may be that something akin to the Hawthorn effect is created in the one-to-one situation. The effect of the question form is an interesting one and of considerable importance to this project. It is returned to when discussing the actual error analysis used for this study.

Error analysis adopted for this study. For the purpose of this study the error analysis was changed in the following manner.

The first two categories were changed to form the following categories:

1. Reading - identifying the relevant information; and
2. Reading - graphics

Understanding of individual words and symbols was assumed as they were all fairly simple. The inclusion of a section of graphics was considered essential as the questions were not all self contained but required the children to refer to figures outwith the text of the problem. In a number of cases there were illustrations that were not needed for the solution of the problem but that were nonetheless used by the children.

3. Identifying the Operation.

Newman's third category is named "Transformation". This is in some ways an ambiguous category. The subtitle refers to the pupil being able to *select* the correct mathematical process. For the purpose of this study it was felt that to insert a category named "Identifying the Operation" would be more useful. It is considered that the pupil needs to identify the correct operation prior to being able to transform it into a suitable symbolic form for carrying out the operation. For example, a pupil may recognise that a problem requires subtraction for its solution but still not be able to transform it correctly. This happened in several cases where the child would proceed to put the smaller of the two numbers to be subtracted at the top of a column subtraction sum. Thus it is suggested that the identification process precedes the transformation process.

4. Transforming.

This category is retained and refers to the child's ability to produce a suitable representation for solution.

5. Process skills.

This was Newman's fourth category. It contained a number of subcategories for arithmetic skills. For the purpose of this project the child has been labelled as deficient in process skills if s/he was not able to carry out the required operation once the process had been identified and the problem correctly transformed.

6. Encoding.

This has been retained as a category; however, the answer was considered correct if the mathematical operation had been successfully completed.

Omitted categories. The categories of "Carelessness" and "Motivation" were not considered useful in their own right for this study for reasons already suggested by Watson.

Nor was "Question form" included. However, it was clear that the way a question was presented had an effect on the child's ability to cope with it. The difficulty of the words and the symbols are not the only parts of the problem which may cause difficulty. The way the essential information is set out - that is, whether it is within the actual question or elsewhere - can also cause problems: it is sometimes stated above a collection of problems, so that the child is required to read from more than one place. In this project the first two categories: Reading - identifying the relevant information; and Reading - Graphics, were considered to cover this omitted category. However, further refinement in this area were found necessary, these are discussed in Chapter 5. These changes were dependent on the responses received. They covered not only the language within the problem to be solved but any other language on the textbook page, such as headings, which the child may choose to use.

The error analysis used for this study thus consisted of six categories as follows:

1. Reading - identifying relevant information
2. Reading - graphics
3. Identifying the operation
4. Transforming
5. Process skills
6. Encoding

These were used for displaying the results.

4.4 Results and Discussion.

The results are contained in two tables. Table I shows the number of children successfully completing the question compared to those that did not. Table II displays the types of difficulties that were experienced by the children for each problem.

TABLE I

Number of children successfully completing the interview problems compared to those not successful.

Problem No.		Successful**	Unsuccessful**
1.	(add)	5	9
2.	(sub)	4	10
3.	(add)*	4	9
4.	(add)	8	6
5.	(sub)	2	12
6.	(sub)	2	12
7.	(sub)*	4	9
8.	(sub)*	6	7
9.	(sub/add)*	0	13
10.	(multi)	10	4
11.	(multi)	4	10
12.	(multi)*	6	7
13.	(multi)*	10	3
14.	(div)	3	11
15.	(div)*	7	6
16.	(div)*	6	7

Total number of children participating: 14

* One child with severe learning problems was not able to attempt all the problems and has thus been omitted from the analysis of these problems.

** A child was considered successful if he/she was able to complete the problem without any outside prompts or assistance.

4.4.1 Factors affecting problem difficulty. It can be seen from the table above that there was a considerable variation in the ability of the children to find solutions to the word problems. Two factors that may have had an effect will be considered here: arithmetic operation and number of arithmetic operations required. Examination of SPMG textbook Stage 2 and the accompanying teachers' handbook suggests that, for the textbook writer and the teacher, arithmetic operation is considered the main factor determining problem difficulty.

(i) Arithmetic operation. Taking an average of the successful responses over the same problem types suggests subtraction to be the most difficult with an average of 3.6 correct responses, division and addition fairly close together with 5.3 and 5.7 respectively and multiplication the easiest with average of 7.5 correct responses.

This is, however, not a particularly useful comparison at this stage as no attempt had been made to equate the problems in relation to other factors that may affect difficulty . Two such factors might be: (a) most recently learnt arithmetic operation and, linked to this (b) the size of numbers used in recently learnt operation.

(a) The most recently learnt and practised operation may be the most easily accessed from memory and could therefore prove to be the easiest to execute.

(b) When children are taught new concepts, such as multiplication or division, the numbers involved are usually small. Early multiplication and division normally employ only one number above ten, with the other number below ten.

Thus the size of the numbers used to practice the different arithmetic operations at the Primary 4 stage of learning varies depending on the stage in the learning process that has been reached. As subtraction is learnt earlier than either multiplication or division, number size may affect subtraction to make it considerably more difficult at this stage.

Size of number acting as a clue to operation. It is interesting to note also that the size of the numbers involved may act as a clue to the operation. This was shown by one child. During the interview he identified a problem as one requiring division. On reflection he changed his mind because he recognised that he had never had to divide with numbers above ten. This is an interesting insight in that it suggests that some children use any available clue in order to solve the problem in a fashion that is consistent with their logical understanding of the situation. This type of behaviour was also shown by another child who selected division as the operation when multiplication was in fact correct. She had mis-read the heading for the page "revision" and instead interpreted it as "division". Chapter 2 and Chapter 6 both suggest that schema theory is useful in explaining children's mathematical behaviour. These two incidents lend further support for such a theory.

(ii) Number of arithmetic operations required. This seems to be a useful indicator to problem difficulty. The only problem containing a two step solution requiring two different operations proved the most difficult. However, it must be considered that one test item does not provide sufficient evidence to suggest that multi step problems are more problematic than single step ones. It does have support from other research. The finding that an increasing number of "steps" (that is the number of mathematical operations required for the solution of a problem) increases the difficulty is in keeping with Nesher (op.cit.) and the Stanford research group quoted in Nesher's research.

The above discussion has focused on the arithmetic aspect of the word problem. It does not provide a complete investigation of the factors that affect problem difficulty. The error analysis provides further breakdown of the difficulties experienced. This highlights the fact that some word problems are found difficult at the comprehension level rather than the procedural level. Table II shows these difficulties in relation to the individual problems.

Table II

Types of difficulties encountered by the children in the study

Problem	Cat. 1 Reading: identifying rel. info.	Cat. 2 Reading: graphics	Cat. 3 Identifying operation	Cat. 4 Trans-forming	Cat. 5 Process skills	Cat. 6 Encoding
1. (add)	0	4	1	0	5	0
2. (sub)	3	0	7	4	8	0
3. (add)*	8	0	3	0	3	0
4. (add)	0	0	4	1	4	0
5. (sub)	0	1	7	8	7	0
6. (sub)	7	1	3	5	11	0
7. (sub)*	0	0	3	2	9	0
8. (sub)*	0	0	0	0	7	0
9. (sub/add)*	5	0	6	9	11	0
10. (multi)	1	0	2	3	3	0
11. (multi)	3	0	1	2	7	0
12. (multi)*	6	2	4	3	2	0
13. (multi)*	0	0	0	1	2	0
14. (div)	0	0	6	4	9	0
15. (div)*	0	0	2	1	5	0
16. (div)*	3	3	2	1	6	0

* Problems not completed by the child with learning difficulties.

Note: As the children were helped towards a solution if they could not manage on their own, they may have had difficulties in more than one category.

4.4.2 Difficulties evidenced by the error classification.

Categories 1 and 2. Difficulties in these two categories seem to be particularly apparent in problems 1, 3, 6 and 12. Examination of the actual problems show that difficulties in these two categories may relate to two types of difficulties: (i) to problem presentation difficulties, i.e. the way the problem is presented on the page, and (ii) graphics.

(i) Problem presentation difficulties. An analysis of the problems mentioned above show that they required the children to find the essential information from outwith the actual problem text. Some of these problems were also what Shuard & Rothery (op.cit.) refer to as stem questions. These are subdivided questions which may contain all the essential information at the top which requires the child to backtrack to obtain information for subsequent parts of the question.

(ii) Graphics. Graphics was in evidence only in a few of these problems. Where they were used difficulties were in evidence. The seemingly simple first problem read:

> Here are the marks given to a
> skater by the judges.
> Find the total mark.

Below this statement a picture showed six judges holding up the marks. These marks were out of ten. Many of the children proceeded by reading two single digits together to form a two digit number, when in fact only a single digit was intended. It is possible that the children were so used to adding two digit figures, using the standard algorithm, and that they did not therefore expect to have to add single digit ones at this stage of their learning. It was also clear from questioning the children that those who performed correctly had some knowledge of the context within which the problem was set.

Categories 3 and 4. Identifying the operation and transforming it created more difficulties than did the previous two categories.

(i) Identifying the operation depends on understanding the semantic relationships contained within the problem. However, in some cases the semantics may be bypassed with the child relying on a particular cue, either within the problem or on the textbook page, to suggest the operation. Interestingly enough, one of the skilled participants seemed to be particularly reliant on external cues such as headings. Her main problem occurred in the only two-step problem. Here subtraction and addition were required, in that order. The heading told her that she needed both operations but she was not very sure about how to proceed. The other high performing girl relied on no external cues and was able to extract the information she needed from the problem statement. She made few mistakes and these were probably mainly due to carelessness.

(ii) Transforming difficulties were particularly evident in the case of subtraction where the correct format of the standard algorithm seemed to be poorly understood.

Categories 5 and 6. Process skills and encoding.

(i) Process skills were lacking, particularly in the subtraction algorithm. This links to the difficulties experienced in transforming these types of problems. However, it must be noted that this category included mistakes ranging from slight slips, which were corrected after a minor prompt, to an extremely poor grasp of number bonds.

(ii) The encoding category (writing the answer in an acceptable form) has been retained although it showed no error. With help at the process stage all the children managed to arrive at the correct solution. The instructions given to the children did not emphasise the need for the answer to include anything more than their solution to the mathematical operation. It was, however, felt that this category should remain and in a further study the instructions given to the children should include an emphasis on the need for the answer to be complete. A complete answer including any necessary wording from the problem would indicate that the child had understood the underlying semantics of the problem. Inability to complete the answer could suggest that the child has relied on a verbal cue within the problem to suggest operation and that the underlying semantics has been bypassed. The effect of verbal cues has been studied by Nesher and Teubal (op.cit.). Of relevance here is their suggestion that a word problem has three levels:

(a) the verbal formulation

(b) the underlying mathematical relationship

(c) the symbolic mathematical expression

Take, for example, the problem:

"In a game of darts Billy King had scored 187 and Jock Scott 223. What is the difference between these scores?"

The verbal formulation is the problem statement. The underlying mathematical relationship is that the difference between two quantities is sought. This can be done in a number of ways. The standard one for this type of problem would be 223 - 187, and this forms the symbolic mathematical expression. However, it is possible for a child to arrive at the symbolic mathematical expression without understanding the underlying mathematical relationship. A child that has learnt by rote the rules (i) "difference between" means subtraction; (ii) always put the bigger number on top, could produce the above symbolic mathematical expression without fully understanding why. That this indeed does happen was suggested by interview responses. Several children when asked why they had used subtraction to solve the problem shown above, responded with "because the teacher says that difference between means take away". This ties in with the argument put forward by Nesher and Teubal. They suggest that the problem solver may bypass level (b) if the problem contains a verbal cue that suggests the mathematical expression. If a child is not able to refer back to the problem and suitably encode the answer it could be suggested that s/he had relied on a verbal cue and has not properly understood the relationships involved.

Conclusion - types of difficulties encountered.

The above analysis shows that there are a number of factors that influence problem difficulty. Some of these factors may be part of the actual problem statement. Other factors are those that relate to the layout of the problem and the use of graphics.

4.4.3 Sex differences in the types of difficulties experienced.

Some of the research in this area has indicated no significant sex differences in problem solving ability (e.g. Linville, op.cit.). Others such as Marshall (1983) reported evidence of such differences. She comments in her study that most of the research in this area has either used total test scores or rates of success on particular items. The findings when using total test scores suggest some support for sex differences in problem solving but this is not supported by all the studies conducted. Looking at the interaction between test item and sex the findings are more positive in supporting the existence of sex differences. Marshall's study uses faulty responses from multiple choice tests to investigate whether there is a sex difference in the type of faulty responses chosen. This she suggests is confirmed by her findings.

Evidence from the individual results of this study suggests sex differences in the type of difficulties that were experienced.

	Girls	Boys
Conceptual errors	56	43
Process skills	38	62

The first three categories - Reading: identifying the relevant information, Reading: graphics, and Identifying the operation - are considered conceptual. They amount, perhaps, to the comprehension of the question.

A chi-squared test applied to these results suggests a small but significant difference ($\chi^2=6.16$, DF=1, $p<0.05$). It must be noted, here, that the relationship between the two groups is slightly distorted with less boys in the sample for seven out of the sixteen problems. The sample is not large so it may not be representative. Intuitively there seemed to be a difference between the sexes in that the girls on the whole, though with one or two notable exceptions, were more careful in their calculations. This evidence of a sex difference had not been expected, and was at this stage considered worthy of further study. However, analysis of the results for Phase 1 showed that these differences between the sexes were not in evidence in the larger sample of that phase.

4.4.4 Feasibility of the chosen research methods.

The main aim of this exploratory study was to explore the feasibility of using the task-based interview, in a school setting, as a tool for exploring children's understanding of word problems in a textbook. It is suggested that data above supports the feasibility of this approach in such a study, but that a certain number of changes/additions to the procedure employed should be made before proceeding with further studies. Changes and additions were implemented in the following areas:

(i) Interview schedule
(ii) The use of concrete materials
(iii) Error analysis categories

(i) The instructions for the actual interview were further standardised so that the interview had a set number of questions that were adhered to. Allowances were made, however, so that each child's individual questions were dealt with adequately. The main advantage of the task-based interview is that individual responses can be probed. Inflexibility in the interview procedure would jeopardise this advantage. However, flexibility was achieved by following a set number of questions with prompts allowed for each question. The Newman study (Clements, 1980) used the following questions:

1. Please read the question to me. If you don't know a word leave it out.
2. Tell me what the question is asking you to do.
3. Tell me how you are going to find the answer.
4. Show me what to do to get the answer. Tell me what you are doing as you work.
5. Now write down the answer to the question.

Watson (op.cit.) found that with younger children these questions were not very helpful as they found it difficult to verbalise their thoughts. Experience gained in the exploratory study shows this to be the case with many of the children in the age group used for this study. Hence a simplified format was used. This is described fully in Chapter 5, Section 5.2.6 and will not be reproduced here.

(ii) The use of concrete materials. Another change from the exploratory study was made. This involved the use of concrete materials by the children taking part in the interviews. SPMG encourages the use of concrete materials, and they are available in many classrooms. It was therefore felt that they should be made available in project of this type which aims to operate in the classroom setting. The children were informed at beginning of the interview that the materials were available for them to use as they desired.

(iii) Error analysis changes. The error analysis was adapted from Newman's (Watson, op.cit.). When considering the responses this adaptation of the error analysis was found to be wanting in certain respects: the two first categories relating to reading, and the category of process skills. These were both further refined.

Reading and comprehension. Newman used these two terms, this exploratory study utilised only two subdivided reading categories. It was clear, from the behaviour of some children that they gave an appearance of being able to read correctly. However, subsequent behaviour showed evidence of lack of comprehension. The error analysis was therefore changed to include the use of the first two categories of Newman's error analysis: Reading and Comprehension with their respective subcategories. Newman, however, had no category for Graphics and it was felt essential, due to the type of presentation used in SPMG Stage 2 of the word problems, to retain this category.

Process Skills changes. It was felt that this category required refinement to differentiate between those children making a slight slip and those totally unable to carry out the operation. Thus Newman's subcategories: faulty computation, random response and no response will be used.

To this was added the subcategory "careless error". It was suggested earlier that the main category "Carelessness" was difficult to assess. There are cases, though, when a child's computation is slightly wrong and the child when asked "what did you do there" immediately spots the error and corrects it. This subdivision allowed for differentiation between the child who makes a slight slip and the child who shows a more serious lack of understanding of process skills. It does not make the assumption made in the Newman study, that if a child can do it in the interview situation then he can also do it in the test situation.

One further change was proposed for the task-based interviews: at the outset of the exploratory study it was felt important that each child was helped to complete successfully the problems they were asked to attempt. It was felt at the completion of the exploratory study that it was questionable if this was in fact necessary. Some of the problems were too difficult for some of the children. It may be better not to struggle on when incomprehension reigns. Thus it was intended in the main study not to aid the children, except for explaining words if asked to do so (see Interview Question 1). This proposed change was only tried out for the first two interviews in the main study. It was found to be an unsatisfactory way to progress as it did not allow children to show if they possessed procedural skills even if they lacked conceptual ones. It was thus abandoned. This change is further discussed in Chapter 5, Section 5.2.6.

Link between language and mathematical difficulties. A subsidiary aim of this study, mentioned in the introduction, was to investigate the possibility of identifying pupils who may be experiencing difficulties in mathematics due to language difficulties. It was suggested in the previous chapter (Section 3.3.2) that children scoring below average on a language test but above average on a computation test may be having difficulties that are language rather than mathematically based. However, examination of the results of the computational part of the mathematics test and the results of the language tests, showed that the possibility of identifying such pupils was unrealistic. Only two out of the fourteen could in any way be considered as falling into such a group. Investigation along these lines was therefore not continued.

4.6 Conclusion.

This exploratory study set out to test the feasibility of gaining insights into children's problem solving behaviour by conducting task-based interviews in a classroom setting. An error analysis, based on that used by Newman, was used to categorise the data thus collected. The exploratory study showed that it was indeed feasible to collect useful data in this manner. It also showed that difficulties relating to recently taught classroom material is not limited to a small number of children. These types of difficulties were also shown by children who had not been identified by their teachers as having learning difficulties.

However, it was felt that the interview procedure needed standardising and a standard format was suggested. Changes to the adapted error analysis were also proposed. This it was felt would improve the discrimination provided by this type of analysis.

To develop further the understanding gained through this study a number of further studies were conducted. These form the contents of chapters 5, 6 and 7.

1. Chapter 5 examines the types of difficulties experienced by other children in this age group. The error analysis suggested that word problems created difficulties not only for children considered as having learning difficulties but also within the middle ability range. Children of "average" ability were therefore chosen as providing a worthwhile focus for this study. This sample included pupils from a number of school. This provided a wider understanding of the types of difficulties experienced by these type of children. This further study forms Phase 1 of the main study.

2. Chapter 6 examines more closely the problem variables. The rewritten versions of some of the problems used in Phase 1 were presented to the Phase 1 sample. This allowed for further consideration of the location of the

actual problem difficulty. This forms Phase 2 of the main study.

3. Chapter 7 looks at the presentation of the original problems alongside the rewritten versions. The presentation of the reconstructed problems in Phase 2 without the original problems presented difficulties in assessing the effects of the structural alterations. A further phase - Phase 3 - was therefore conducted. This presented the rewritten problems alongside the original ones to a new but similar sample.

The exploratory study thus provided the understanding required for a focused investigation of the types of difficulties that some children experience when trying to solve word problems.

CHAPTER FIVE – MAIN STUDY: PHASE 1

5.1 Introduction.

The previous chapter looked at the exploratory study carried out. This chapter will describe the first phase of the main study. The intention of this research was to explore and classify the difficulties encountered by Primary 4 children when solving word problems in SPMG textbook Stage 2. The exploratory study examined the feasibility of using the task-based interview as the main instrument for data collection. To analyse the data collected an error analysis was created based on Watson's error analysis (op.cit.). Watson suggested that the interview technique coupled with his error analysis could provide teachers with a more effective method for exploring their pupils' difficulties.

It is hoped that this research will prove useful to teachers as it looks closely at a number of problems from the commonly used SPMG textbook, describes in detail the difficulties experienced by the children trying to solve them, and provides a tool for analysing the difficulties the children experience. The results from the exploratory study suggested that the task based interview coupled with the error analysis provided useful data on children's understanding of mathematics. It was, however, suggested that the main study should differ from the exploratory study in the following way:

(i) that the focus was narrowed to study only the "average" pupil. This decision was made as the exploratory study indicated that these pupils provided a population well worth studying, as the number of difficulties they encountered were greater than expected.

(ii) that this type of pupil was studied in a number of different schools. The exploratory study was limited to one class. It may be that the type of difficulties experienced by the pupils in this class were not representative of a wider population.

(iii) that the number of problems used was limited to thirteen. The problems the children were asked to do were on pages already covered by the teacher. As the main study was carried out in the period October to December the concept of division had not been taught to all the classes. Thus the ~~two~~ three final division problems were left out.

(iv) that the revisions suggested by the exploratory study in relation to the interview technique and the error analysis were put into effect.

This chapter will follow a format similar to the preceding one. The section on method will also contain an explanation of interview technique and error analysis changes. This will be followed by the results and a discussion of these results. As the data is gathered from a number of schools it would be useful first to present the overall picture of the total population. Differences and similarities between the schools can then be highlighted and discussed. This will be followed by a more detailed discussion of some of the problems used. Individual pupil profiles will then be discussed. A number of issues arose out the results of this phase of the research which potentially have a bearing on the teaching of mathematics. These points will be raised in Section 5.4 - "Issues Arising". Section 5.5 "Further Research" will set the scene for Phase 2 of this study, and this will be followed by a summing up conclusion. The points to be looked at in the further discussion are as follows:

- The "average" pupil was identified by using a standardised mathematics test. The question of the extent to which one can speak of an "average" pupil will be raised.

- SPMG publishes a handbook for teachers: to what extent do the suggestions for teaching of the particular problems used relate to the difficulties the children appeared to encounter?

- The textbook expects problems to be represented using the relevant standard algorithm, and seems to assume that a correct answer implies understanding of that algorithm. For example, a multiplication problem is expected to be carried out as multiplication, not repeated addition. That this kind of expectation exists is shown by some of the page headings in the textbook, the way children's work is recorded and in the instructions in the teachers' handbook. It seems that some children may use idiosyncratic methods successful at a simple level, but which may be ineffective at a slightly higher level, and that there is a danger that the structure of the materials/teaching may allow this to pass unnoticed.

- The exploratory study suggested sex differences in the type of difficulties experienced by boys and girls. Do these differences also manifest themselves in the more homogenous population in this phase of the main study?

- What effect does the availability of concrete materials seem to have on the children's ability to solve the word problems? The use of concrete materials is not the focus of this study. However, their availability to the children participating in this project provided interesting insights into how the children used these materials. SPMG and other educationalists advocate their use, possibly without a thorough survey of how children use them in the classroom. It was therefore considered useful to include the observations on the use of concrete materials that were made during this phase of the project.

5.2 Method.

5.2.1 Design. The exploratory study looked at the difficulties faced by children from the whole ability range. The purpose of this stage of the study was to focus on the "average" child. In order to identify a group of forty average children from five different classes, a standardised mathematics test was used. These children were then interviewed individually to access each child's understanding of thirteen word problems from SPMG textbook Stage 2. These interviews were taped and were later analysed together with any workings the children had produced whilst trying to solve the problem. An error analysis was then carried out on this data. This analysis has been described in the previous chapter as has the interview method. The changes that were made will be explained in Sections 5.2.6 and 5.2.7.

5.2.2 Materials. (a) SPMG textbook, paper and pencil

(b) concrete material

(c) mathematics test

(a) The thirteen word problems used are shown in Appendix A. The problems used are the first thirteen in this appendix.

(b) Dienes wooden units, tens and hundreds were available for the childrens' use.

(c) The mathematics test used was the standardised Y1 series by D. Young (1979), intended for an age range 7:5 to 8:10. As the majority of the children fell within this age range it was considered acceptable. The time allowed for the test was forty minutes. This is rather long for this age range but there was no suitable shorter test. It was evident from the restless behaviour that some of the children exhibited towards the end of the test that it was too lengthy. The test consisted of an oral section, a section with computation, and one on word problems. It has been criticized by Jim Ridgway (op.cit.) for not being a particularly good test on a child's all round mathematical ability. This is possibly a fair criticism but finding an ideal mathematics test that is also relatively simple to administer and not too expensive or time consuming is difficult. If it was to be a good all round test for this age level it would seem important to include a practical element. However, for the purpose of this research this would have been impossibly time consuming. A lengthier, possibly more wide ranging test will run into other difficulties - one such is the lack of concentration that many children in this age group exhibit. This test provided enough evidence of the children's mathematical ability to give a suitable population for this study. The limitations of test data will be commented on when looking at the concept of the "average" pupil.

In Chapter 3 (p.78) it was suggested that verbal problems should be omitted from this standardised assessment, as the language variable was the one under study. However, the exploratory study (Chapter 4) suggested that a clear cut language/computational abilities dichotomy is difficult to find. A combination of factors seem to affect the difficulty of the problem, of which language is one. Thus this study was not looking solely for pupils who seemed to be detrimentally affected in their mathematics due to a language deficit. It was therefore no longer considered essential to use a mathematics test that did not include word problems.

5.2.3 School/classroom backgrounds. Each school is dealt with individually as there were quite distinct differences both between the schools and the classroom practices. The schools are numbered 1 to 5 and referred to by this number. All classes involved are Primary 4. The ages of the pupils involved in the study ranged from 7:9 to 8:8 at the date of the mathematics test.

School 1. This is an inner city school, albeit in a small city. The only outlook is on to other houses and the school is surrounded by streets on all four sides. The catchment area is predominantly working class. The teacher in this class kept firm but kind discipline and insisted on a high standard of work. She felt that language was an important part of teaching mathematics. She had no groupings in mathematics but taught the whole class together by topic. So, for example, when multiplication was started this was first "class-taught" for approximately a week and then followed up by work as appropriate for each individual in the workbook and the textbook. There were four children in the class who had not yet reached the standard required for Primary 4 and these children, although joining with the whole class lesson as appropriate, were given different work based mainly on SPMG material aimed at Primary 2 and 3. The textbook was not followed page by page but rather according to topic. There were 26 pupils in the class and the total school roll was 166.

School 2. This is a suburban school with a mainly middle class catchment area with an overall roll of 200. The school has very little space both outside and inside. The class is divided into three ability groups for mathematics. The teacher commented on the fact that these groupings were "inherited" from the previous Primary 3 class teacher and that he was not totally happy with them. Out of the pupils selected to participate in the research, one was in the top maths group in the class, four were in the middle group and three were in the bottom group. This indicates a discrepancy between the results of the mathematics test and the teacher's groupings. Maybe this to some extent confirms his dissatisfaction with the inherited groupings. He felt that a number of pupils in the top ability group should possibly be moved to the middle group. However, nothing was done to make these changes during the period of that this research took place. The workbooks and textbook of SPMG is worked through roughly in the order suggested by the layout of the material. The class was large - 33 children - of which about twenty were boys. The classroom was cramped and possibly therefore gave an appearance of being untidy. There was quite a lot of noise in the room, some of it caused by the lack of space for movement.

School 3. This was another city school with a very depressed catchment area which reflected itself in the school. The school buildings themselves were large and old fashioned but quite roomy. Only about half the class had reached SPMG Stage 2 at the start of the study. It was, according to the teacher, a very unruly class. She seemed to be firmly in control, though possibly, (and this was her own suggestion), at the expense of not getting on with the teaching. One distinct problem that seemed to cause a great deal of trouble in this class was the lack of pencils. The teacher complained that the whole years supply was already used up and that no more were available. Few of the children brought their own. One child was observed carefully sharpening his pencil, then breaking it and starting the whole process again. This behaviour went on until it was nearly playtime. When asked by the teacher why he had not completed his work he complained that his pencil was broken. With 24 pupils in the class and several of them constantly complaining of lack of a pencil it created a constant nuisance. The range in mathematical ability was wide, the child with the highest overall score on the standardised test was found in this class, but generally the standards were not high. Apart from one child who was in the middle ability group the rest of the children that were part of this research all came from the top ability group. The poorest child in this class was still working on infant mathematics. He required everything to be read to him and was unable to progress, on his own, with anything that required the reading even of the simplest instruction. The total school roll was 250.

School 4. This was the only Catholic school within the sample. Again, it was an inner city school totally surrounded by streets. The teacher worked to some extent in groups for practical tasks but tended mainly to use SPMG as an individualised scheme and pupils moved on at their own pace. It was a large class and discipline was firm. There were 30 children in the class and the school roll was 215. The headteacher commented favourably on SPMG as long as it was used as a resource and suited to individual teachers' and pupils' needs and not followed too literally, page by page.

School 5. Another inner city school bordered by streets and a large main road. This was a large class of 33 of whom 20 were boys. The high proportion of boys was seen by the headteacher and class teacher as one of reasons for the unruliness of the class. There were a number of children with considerable behaviour difficulties within the class. The peripatetic learning support teacher spent a large amount of her time in this class. The headteacher stated a dislike of SPMG. She felt it contained too much language and emphasised the use of concrete materials unnecessarily. The class worked in ability groups for maths but any new topics were introduced in a whole class lesson. Out of the eight pupils participating in the research 3 came from the top ability group, 4 from the middle group and 1 from the low ability group. The total roll in the school is 185.

5.2.4 Subjects. Eight pupils from each of the five schools were involved in this phase of the study. This gives a total of forty subjects. They were chosen from those that had scored around average for this particular population on the standardised mathematics test. An attempt was made to keep the balance between the sexes. However, two of the classes had a large proportion of boys. This meant that from these classes there were likely to be more boys falling into the average category. This proved to be the case, and from these two classes there were more boys in the sample. This was compensated for by having slightly larger proportion of girls from two of the other schools. Altogether the sample consisted of nineteen boys and twenty-one girls. The age of the subjects ranged from 7:9 to 8:8 at the date of the test. At the time of the interviews the age range was 7:9 to 8:10.

5.2.5 Apparatus. A Sony TC-D3 stereo cassette recorder with a microphone was used.

5.2.6 Procedure. All the work with the children took part during the period after the October holiday break until a week before the end of that same term. All the children were tested and interviewed during the forenoon.

The standardised mathematics test. This was carried out in the fortnight preceding the start of the task based interviews. In each school, the whole class was tested at the same time. A mean was worked out for each school and for the whole sample. The whole sample mean provided the basis for selection of subjects. Only pupils falling within +4 and -4 of the mean for the whole sample were to be included. In one class this only provided six subjects. Two more were included; one with the nearest score above +4, and one with the closest score below - 4.

The task-based interview was used as in the exploratory study. It was suggested in the discussion in chapter 4 that instead of prompting children to continue on a problem or correct a mistake the child would just be asked to go on to the next problem. This method was used for the first few interviews. However, it was felt that this did not reveal as much of the children's understanding as would the interview accompanied by prompts. The procedure was thus changed and this is reflected in the error analysis of the problems, and will be discussed below. The need for such a change is illustrated by an example. One child started to make a subtraction error. No attempt was made to correct this error, but the prompt "can you do that?" was sufficient to cause the child to correct her error and arrive at the correct solution. In other words she had the ability to work the sum correctly but had slight problems in retrieving the essential information to do so. This links to Gelman and Meck's discussion on conceptual, utilisation and procedural competence. As the child was able to identify the necessary operation for the solution of the problem she could be deemed to have conceptual and utilisation competence but was weak, though obviously not completely lacking in procedural competence (Gelman & Meck, 1986, pp. 29-57).

The setting for the interview was broadly similar for all the children. It was possible in each school to find a relatively quiet room away from the rest of the class. All the children were quite happy to take part in the interviews. This was reflected in the children's attitudes to the researcher. Several of them asked if they could do it again!

The use of concrete material was a change from the procedure in the exploratory study. It was felt that some children might find this a help as SPMG stresses the use of concrete materials. The material was shown to the child and the child was told that s/he could use it if s/he felt the need for it.

The interview schedule which was followed (with some deviation when it was necessary to respond to individual children's requirements) is set out below:

Interview Schedule:

1. As you know I have been looking at your maths in the classroom. Now I am wanting to find out how you do word problems, so I am going to ask you to do a few from the textbook. I am going to record what we both say so that I can listen to it again later. All right?

2. (show the child the concrete materials) Do you use these sometimes? If you want to use them for any of the problems just go ahead and use them.

3. Are you ready to start? (switch on tape-recorder).

4. Could you do No .... (if child looks uncertain) Do you know which one it is?

5. (if the child was stuck and did not appear to be taking any action) Can you read it to me?

6. (if complete encoding did not take place) Is that all? (followed by further prompts if necessary, except in the cases where the children were struggling to produce any answer. In these cases the encoding was sometimes left out).

7. (after the completion of a number of problems). Do you read the heading at the top of the page?

5.2.7 Error Analysis. The following categorisation of errors was suggested from the exploratory study:

Category 1:	Reading errors	(i)	words
		(ii)	symbols
		(iii)	graphics
Category 2:	Comprehension errors	(i)	general understanding
		(ii)	specific terms
Category 3:	Identification of operation errors		
Category 4:	Transforming errors		
Category 5:	Process skill errors	(i)	faulty computation
		(ii)	random response
		(iii)	no response
		(iv)	careless slip
Category 6:	Encoding errors		

Two points need to be considered in relation to the error analysis: a) actual and potential changes to error analysis categories that emerged during analysis of the main study data, and b) categories that for the purpose of this study do not constitute failure to reach a solution.

a) **Error analysis revision.** The revised analysis showed itself to lack one subsection in relation to comprehending the material. There was no category corresponding to cases of children not being able to locate all the information they need to solve the problem. This particular difficulty was in evidence especially when this information had to be found outwith the actual word problem (see for example, Problem 3). A third subsection was therefore added to Category 2: Comprehension: (iii) identifying relevant information. A second change that was not made, but should be considered for any future use of this error analysis, was that subsection (iii) of Category 1: Reading - Graphics - was not intended to apply only to the ability to read graphics but intended also to record error in understanding of any graphics that were part of a problem or were taken by the children to be part of a problem. This subsection would therefore be more accurately placed if contained within Category 2 - Comprehension. Here it could form subsection (iv).

b) **Omitted categories.** A child was considered to have successfully completed the problem if the solution was correct and s/he needed no prompts or explanations during the process of solving the problem. However, assistance in Category 1 - Reading of words and symbols [subsection (i) and (ii)] refers only to the child's ability to read aloud or pronounce the words in the word problems, not to his/her ability to comprehend. These two abilities are not the same as the following case illustrates. One child who was unable to read/pronounce the word "altogether" was quite able to solve the problem successfully without any further help once this word had been read to him. Thus, for this stage of the research, it is considered that lack of ability to read the word should not be counted as a difficulty when deciding whether the child has successfully completed the problem without outside aid. It is, however, worth retaining as a category if this categorisation system is to be used by teachers. The fact that seemingly common mathematical words cannot be read by some children points to the dangers of using them in schemes that might be considered suitable for children to work with on their own without too much teacher help. If the teacher does not have sufficient time to help a child with any reading difficulties s/he might experience it is possible that this child could be held back solely due to this reading difficulty.

A second category - Category 6 - Encoding was not taken into account when deciding whether the child had successfully completed the problem or not. The reason for excluding this category was that there was inconsistency in this research when demanding complete encoding. The decision to omit this category at this stage of the research was due to the great difficulties that some children experienced when trying to reach a solution. It was decided not to tax them with further demands by demanding correct encoding. There were also a number of problems where suitable encoding was quite difficult. An example of this is problem 5. It stated: "During a game of darts Billy Smith had scored 187 and Jock Scott 223. What is the difference between these scores?" What is the correct terminology for a game of darts - should it be points? This was not known by many of the children and no indication was given by the problem. It was, however, felt that it should be retained as a category if the error analysis is to be used by teachers. For a teacher it can be a useful indication of whether the child has thoroughly understood a particular problem.

5.3 Results and Discussion.

The results of this study are displayed in three tables:

Table I looks at the broader picture showing the success rates across the problems from the different schools

Table II examines the types of difficulties that the different problems created

Table III shows five individual profiles to indicate the differences that exist amongst a group of pupils identified as "average"

After looking at these results the word problems used are considered and set into the context of the children's curricular experience.

The results shown in Table I allow consideration of the following:

(i) the relative difficulty of a problem, and the effect of the arithmetic operation on the difficulty of a problem.

(ii) similarities and differences between the schools.

TABLE I

Number of children successfully completing each interview task compared to those not successful.

School	1		2		3		4		5		Total	
Problem No	S	US	S	US	S	US	S	US	S	US	S	US
1 (add)	1	7	3	5	3	5	2	5	2	6	11	28
2 (sub)	4	4	3	5	4	4	1	6	2	6	14	25
3 (add)	5	3	1	7	6	2	1	6	1	7	14	25
4 (add)	7	1	4	4	6	2	5	2	3	5	25	14
5 (sub)	3	5	4	4	1	7	2	5	1	7	11	28
6 (sub)	1	7	1	7	2	6	1	6	0	8	5	34
7 (sub)	4	4	5	3	3	5	1	6	0	8	13	26
8 (sub)	4	4	4	4	4	4	4	3	1	7	17	22
9 (sub/add)	3	5	1	7	3	5	0	7	1	7	8	31
10 (multi)	4	4	5	3	4	4	5	2	1	7	19	20
11 (multi)	4	4	3	5	5	3	3	4	0	8	15	24
12 (multi)	5	3	2	6	2	6	0	7	2	6	11	28
13 (multi)	5	3	5	3	4	4	6	1	2	6	22	17
Total	50	54	41	63	47	57	31	60	16	88	185	322

Notes:

Pupils in each school: 8 (except School 4 where the number was 7. A transcript was lost due to technical error. This child is not included in later samples as she left the school)

Difficulties in reading and encoding have not been included when determining whether a child was successful or unsuccessful. The reason for excluding these categories has been explained in Section 5.2.7.

(i) Problem difficulty and effect of arithmetic operation. Problems varied considerably in the amount of difficulty they presented. Only five children managed to solve the most difficult one - problem 6 - whilst the easiest problem - no. 4 - was solved by twenty-five children. In relation to arithmetic operations, problems involving subtraction pose the greatest difficulties but problem 1, which required addition, and problem 12, which required a simple multiplication, also created numerous difficulties. This suggests that arithmetic operation alone does not determine problem difficulty. Chi-squared tests comparing the three arithmetic operations confirm these suggestions. Combining the total responses for the three arithmetic operations show a significant difference between them: $\chi^2 = 7.069$, $p < 0.05$ (DF=2). Further chi-squared tests show that this difference is due to subtraction problems being significantly more difficult than either addition or multiplication. There is no statistically significant difference between the multiplication and the addition problems. However, the notion that arithmetic operation alone is not the determinant of difficulty is confirmed by the comparison of the easiest and most difficult problem within each operation. The relevant data is displayed below:

Addition (problems 1 and 4): $\chi^2 = 8.718$ $p < 0.005$ (DF=1)

Subtraction (problems 6 and 8): $\chi^2 = 7.661$ $p < 0.01$ (DF=1)

Multiplication (problems 12 and 13): $\chi^2 = 5.253$ $p < 0.05$ (DF=1)

Table II, which looks at the different types of difficulties created by the problems, will provide further details of factors that may affect problem difficulties.

(ii) **School differences/similarities.** There were certain significant similarities between schools. The two most difficult problems (nos. 6 and 9) were found difficult by most of the children in all of the schools, thus suggesting an inherent difficulty in the problem. Problem 1, which shared position with problem 5 as the third most difficult, was also found generally difficult by children from all schools.

The mathematics test data provides some information as to the composition of the class in terms of abilities. This data is shown below:

	Mean scores by schools	Standard Deviation
School 1	25.76	11.7
School 2	30.94	8.9
School 3	23.17	11.7
School 4	26.53	11.3
School 5	29.23	10.57
Mean score: whole sample	27.13	

Analysis of the results from Table I compared to those of the mathematics test suggests that teacher variables and other factors make a definite impact on the retention of recently learnt concepts. School 1 which showed the greatest number of successful responses in Table I ranks as the fourth school on the mathematics test. School 5 was the second most successful school on the mathematics test yet performed worst on the task-based interviews.

Chi-squared tests on the total responses from the five schools show a highly significant difference between the schools. Using the data from all the schools a chi-square value of 30.0301, $p < 0.0000$ (DF=4) is found. Closer analysis of the data shows that this significant difference is mainly due to the large number of unsuccessful responses from pupils in School 5. A chi-squared test was carried out comparing School 1 (the school with the most successful responses) with School 4 (the school with the second most unsuccessful responses) to explore further this difference. This shows no statistically significant difference thus supporting the claim that the statistically significant difference was due to the performance of the children in school 5 compared to the rest of the children.

Table Ia extracts the information from Table I according to problem type. This allows for comparisons, using chi-squared tests, of the performances of different schools on the different types of problems.

TABLE Ia: An extract from Table I showing the different responses according to problem type.

School	1		2		3		4		5		Total	
Problem type	S	US	S	US	S	US	S	US	S	US	S	US
Addition	13	11	8	16	15	9	8	13	6	18	50	67
Ratio	1:0.5		1:2		1:0.6		1:1.6		1:3		1:1.3	
Subtraction	16	24	17	23	14	26	9	26	4	36	60	135
Ratio	1:1.5		1:1.4		1:1.9		1:2.9		1:9		1:2.3	
Multi-plication	18	14	15	17	15	17	14	14	5	27	67	89
Ratio	1:0.8		1:1.1		1:1.1		1:1		1:5.4		1:1.3	

Note: This is based on all the problems apart from problem 9 as this problem includes both subtraction and addition.

Chi-squared tests on the data in Table Ia show statistically significant differences between the schools on all problem types:

Addition: $\chi^2 = 9.249$ $p < 0.05$ (DF=4)

Subtraction: $\chi^2 = 13.04$ $p < 0.01$ (DF=4)

Multiplication: $\chi^2 = 13.032$ $p < 0.01$ (DF=4)

Addition. Problems dealing with addition show the smallest statistically significant difference and this difference is mainly due to the difference in responses between schools 1 and 3 and the rest of the schools. A chi-squared test on the data comparing combined scores for schools 1 and 3 with those of schools 2, 4 and 5 supports this. Here a chi-square value of 9.002 (DF=1) gives a $p < 0.005$.

Subtraction. Statistically significant differences between the schools are in evidence for this problem type too. Here, the main reason for the difference is due to the responses from school 5 differing from the rest of the sample. Analysis of the combined scores of schools 1 to 4 compared to school 5 show a greater significant difference ($\chi^2 = 9.001$, $p < 0.005$, DF=1). Comparing the school with the most successful responses (school 2) with the one with the second least successful responses (school 4) gave a non-significant difference. This supports the suggestion that the statistically significant difference found is in relation to school 5 being different from the rest of the sample.

Multiplication. This problem type also shows significant differences between the schools, and, as for subtraction this difference is largely because of the differences in successful responses from school 5 compared to the rest of the sample. Some of the teaching and practising of examples related to multiplication was observed in school 5 and this will be discussed below.

Thus it has been shown by the data above that the responses from School 1 show a generally more successful performance on all problem types on recently taught material than the other schools, in particular in relation to School 5. As the test material suggests this was not necessarily due to a generally higher ability in that sample. It is interesting to note here one particular feature of School 1 that was not in evidence in the other schools: that is the emphasis placed on language understanding by this teacher. She would use language from the SPMG textbook in language teaching and for spelling exercises. Language skills specific to mathematics were also explicitly taught, one example of this is the emphasis on reading and considering each word in a word problem. The children were also expected to be able to explain why they had reached a particular solution. Also worth considering is the fact that this class was not organised in ability groups for mathematics. This is in spite of it showing the greatest equal spread of ability (standard deviation = 11.7) on the standardised mathematics test.

In contrast particularly to school 1, school 5 shows a generally poorer performance on the recently taught material but ranks as the second highest on mathematics test. This may suggest a more ineffective teaching style from their current teacher, with their success in the standardised test stemming possibly from past teaching. School 5 shows up as being especially poor in problems requiring multiplication relative to the other schools. During observation prior to starting the data collection it was noticed that a number of children in this class were getting confused when working individually on textbook multiplication examples. Two digit numbers that were to be multiplied by a single digit number seemed to cause particular confusion when the problem required that a number be carried over to the next column. An example of this would be 2 x 98 giving an answer of 96 (2 times 8 is 16, carry the 1, 2 times 9 is 18 add 1 to give 19, but only the 9 is written and the child continues to carry the 1). This particular difficulty was only observed in this school and may account for some of the problems the children experienced in this area. Another explanation could be that the children in this class were starting to learn division and this may account for some of the confusion. However, if the latter is correct it may be a questionable teaching method that allows the children to move on to a new concept before the previous one is sufficiently well anchored to be remembered a week or two after it has been learnt and practised.

Thus Table I and the related data show that there are a number of problems that seem to create difficulties for a large number of children in several schools and that some problems are particularly difficult for children from specific classes. These most difficult problems may contain certain factors that make them more difficult. Table II looks in more detail at the types of difficulties experienced and may thus throw some more light on the factors that affect the difficulty of a word problem.

TABLE II

Types of difficulties encountered by the children in the study. The numbers refer to the number of children experiencing that particular difficulty.

Problem No	1 (+)	2 (-)	3 (+)	4 (+)	5 (-)	6 (-)	7 (-)	8 (-)	9 (-) (+)	10 (x)	11 (x)	12 (x)	13 (x)
Category													
1. Reading													
(i)	6	2	2	1	6	0	14	9	1	3	1	5	3
(ii)	1	4	1	0	4	4	7	2	6	1	1	9	2
(iii)	7	0	0	0	2	0	0	0	0	0	0	11	2
2. Comprehension													
(i)	9	17	9	6	16	17	10	11	14	11	5	7	3
(ii)	0	0	0	0	0	0	17	7	0	4	0	16	0
(iii)	1	0	22	3	2	11	0	0	15	1	8	24	5
3. Ident. of Oper.	3	15	5	4	15	17	10	12	9	5	2	1	2
4. Transforming	5	5	1	1	13	19	5	2	14	1	1	2	2
5. Process skills													
(i)	18	15	4	6	20	25	11	14	15	10	22	2	14
(ii)	0	0	2	0	0	0	0	1	0	0	1	0	0
(iii)	0	0	0	0	0	0	0	0	0	0	0	2	0
(iv)	7	2	3	2	1	2	6	3	1	0	1	1	0
6. Encode	0	1	1	4	0	0	2	2	0	5	1	1	2

Reading: subsection i: word recognition
ii: symbol recognition
iii: graphics

Comprehension: subsection i: general understanding
ii: specific terms
iii: identifying relevant information

Process skills: subsection i: faulty computation
ii: random response
iii: no response
iv: careless slip

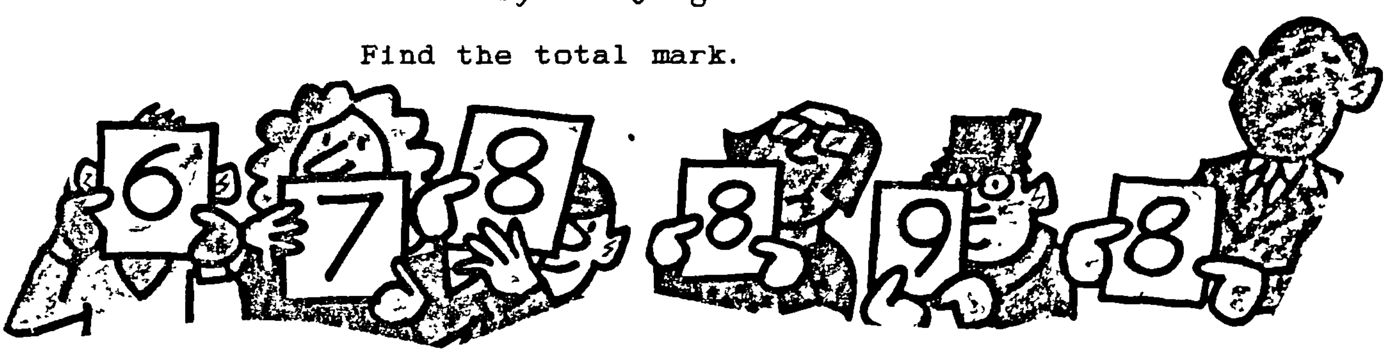
Note: as the children were helped and prompted it may be possible for the same child to have entries in several columns (see Table III - individual profiles)

Number of pupils: 40 (except problems 6-10, nos. = 39)

Factors affecting problem difficulty. Variation in the types of difficulties the children experienced when trying to solve the problems is evident here. Problems no. 5, 6 and 9 caused difficulties in comprehension, transforming and process skills, whilst the difficulties of problem 12 were mainly concentrated in the area of comprehension and graphics. The understanding of graphics also affected the difficulty of problem 1 but here the process skills difficulties were in far greater evidence than in problem 12. It would be useful here to examine each of these five problems in an attempt to locate more specifically where the difficulties may lie. The problems will be reproduced - the graphics will only be shown when it is deemed to have had a significant effect on the children's ability to solve a problem. These problems will then be discussed not only in relation to Table II but also drawing on the individual transcripts as required to highlight particular points.

Discussion of individual problems.

Problem 1: Here are the marks given to a skater by the judges. Find the total mark.



Difficulties in interpreting graphics. Seven children experienced difficulties in interpreting the graphics. The most common difficulty was not understanding that the people pictured each represented a judge and that each gave marks out of ten. Thus the children tended to write the numbers together to form either two digit numbers and create sum such as 89+86+76 or three digit number to create sum such as 898+678.

Comprehension difficulties. A number of children experienced difficulties in this area. It is possible that those identified as having comprehension difficulties were also confused by the graphics. Their interview responses suggested that they had no contextual knowledge about ice skating competitions and thus failed to comprehend the question statement and link it to the information provided by the graphics.

Process skill difficulties. There were a large number of difficulties in this area. A string of digits had to be added and it requires the child to keep a running total. There was very little evidence of the children using some kind of strategy, such as regrouping, for example, to ease this memory difficulty.

Problems 5 and 6: (subsection a is problem 5; subsection b is problem 6)

> During a game of darts Billy King had scored 187 and Jock Scott 223.
>
> a) What is the difference between these scores?
>
> b) How many more does Billy need to make 301?

Comprehension difficulties. The language of subtraction has long been considered difficult (see e.g. E362 Developing Mathematical Thinking, O.U., 1982). This difficulty was in evidence in both these word problems. Many children failed to understand "the difference between" in a mathematical sense. This was evident by responses such as "one is bigger/higher". Even amongst some of the children who identified this word problem as requiring subtraction there seemed to be uncertainty. When questioned as to why they had subtracted they would answer that it was because "the teacher says difference between means take away". A number of these children proceeded by interpreting the phrase "how many more" (in problem 6) as requiring addition. This suggests incorrect use of a verbal cue (Nesher & Teubal, 1976) and that the children are learning to bypass the underlying structure of the problem and only responding at a surface level. A number of the children who experienced difficulties with the comprehension aspect of the problem also found the transforming and process aspects difficult suggesting a general poor understanding of subtraction. However, some were quite able to solve the problem once it had been comprehended.

Transforming difficulties. The standard subtraction algorithm requires the larger number to be placed above the smaller number. This created difficulties for many. Nearly half of the children had this type of difficulty with problem 6 and more than a quarter with problem 5. The responses of some of the children, when they discovered that the final hundred digit was not sufficient for the sum to be completed, were interesting. A few children simply added another digit in front to make the calculations possible, others just ignored it and gave their answer without finishing the sum. When it was pointed out to the children that it was not really possible to do the sum that way, some of them responded by changing it to the correct representation. Others needed further prompting and some required the interviewer to provide the transformation.

Process skill difficulties. Difficulties in this area were numerous. Carrying was found to be particularly difficult, and a number of the "bugs" identified by Van Lehn (1986) were in evidence. Examples of these were "borrow/across/zero", that is if the digit in the column that should be borrowed from is zero the borrowing is done from the next digit to the left; and "smaller from larger", that is when the child subtracts the smaller number from the larger number regardless of the position of this number. It was interesting to note that a number of children who comprehended the problem used their own method for solving the problem. They added on from the smaller to the larger number. This operation was invariably carried out as mental arithmetic. When asked to subtract using the formal algorithm these children had great difficulties. School 2 (see Table I) showed four of the children as successful on problem 5, out of these four, three added on to achieve the correct solution. In this case, successful solution could not be considered as an indication of an understanding of the formal subtraction algorithm. This will be discussed below in the section on problem representation.

Problem 9: Farmer Till had 210 sheep. At the market he sold 88 and bought 25. How many sheep has he now?

Comprehension difficulties. The main difficulty lay in understanding that this problem required two arithmetic operations for its solution. Identifying the relevant information for these operations created a great deal of problem. There are a number of ways in which the correct answer to this sum could be achieved:

1.
```
 210
 +25
 235
 -88
 147
```

2.
```
 210
 -88
 122
 +25
 147
```

3.
```
  88
 -25
  63

 210
 -63
 147
```

However, in reality the majority of the children attempted to achieve a sum like the second of the three above. This directly represents the problem statement. The fact that the sum of the first operation was required before the second sum could be set up confused many. Thus a number of children would correctly attempt to subtract 88 from 210 but instead of using the answer from that calculation they then proceeded to add 25 to 210 and give this as their answer.

Transforming difficulties. This is related to the difficulty experienced in the comprehension of the problem. After some discussion of what the problem required a number of children transformed the the word problem into:

```
 210
 -88
 +25
```

In a sense this was correct but only one child managed to actually carry out this sum and achieve the correct solution. Some started subtracting but found this impossible with three numbers.

Process skill difficulties. Two types of difficulties were in evidence: firstly that a two digit number was being subtracted from a three-digit number and these numbers were placed incorrectly to give an incorrect answer; and the difficulties already mentioned in relation to carrying (see problems 5 and 6).

Problem 12: In a garden there are

2 clumps each with 145 snowdrops,
3 beds each with 72 tulips, and
5 beds each with 50 daffodils.

a) How many snowdrops are there?

Comprehension difficulties. Many children failed to read the essential first part of the problem. Their attention was instead focused on the the flowers drawn around the problem as an illustration. These were counted by a quarter of the children and offered as an answer. When this was questioned and the children were directed to the problem statement above many simply answered with 145. When reading the problem aloud many children ignored the "2" at the beginning of line 2 and the "each". The concept of a "clump of snowdrops" did not seem to have any meaning to many of the children. Once that had been discussed and in some cases shown with a drawing the rest of the problem caused little difficulty.

Process skill difficulties. There were not many difficulties here though it must be stated that a number of children used repeated addition rather than multiplication. Thus another instance of the operation intended by the writers of the textbook not being used and practised as intended.

This discussion has only looked at a few of the word problems in detail, but it does show how an analysis of this nature can help to pinpoint more specifically what aspects of the problems are causing difficulties. The word problems used were chosen at random. In general, the subtraction problems caused the greatest difficulties; however, arithmetic operation alone cannot be deemed to be the determining factor. A word problem with graphics created two types of difficulties: the first when it had to be correctly interpreted and used for the problem solution; and the second when it was not required but nonetheless used. Evidence of the first type is found in problem 1 where the illustration gave the numbers that have to be added. The second type of difficulty is presented by problem 12. Here lack of contextual understanding seems to be a factor to be considered. Phase 2 of this study will look further at six of the problems in attempt to pinpoint more specifically the difficulties located within the word problems. The intention of using the error analysis is not only to locate difficulties within the problems but also to provide profiles of individual children to show how individual strengths and weaknesses can be highlighted. Profiles of five children are presented in Table III.

TABLE III

Individual profiles of five children (one from each school)

III a - Script 7 Mathematics test score: 28

Problem No.	Category 1 Reading i	ii	iii	Category 2 Comprehension i	ii	iii	Category 3 Ident. of operation	Category 4 Trans-forming	Category 5 Process skills i	ii	iii	iv
1 +	-	-	-	-	-	-	-	-	-	-	-	X
2 -	-	-	-	-	-	-	-	-	X	-	-	-
3 +	-	-	-	-	-	-	-	-	-	P	-	-
4 +	-	-	-	X	-	-	X	-	-	-	-	X
5 -	-	-	-	-	-	-	-	-	X	-	-	-
6 -	-	-	-	-	-	-	-	-	X	-	-	-
7 -	-	-	-	-	-	-	-	-	X	-	-	-
8 -	-	-	-	-	-	-	-	-	X	-	-	-
9 -+	-	-	-	R	-	R	P	-	R	-	-	P
10 x	-	-	-	-	-	-	-	-	X	-	-	-
11 x	-	-	-	-	-	-	-	-	X	-	-	-
12 x	-	-	X	X	-	X	-	-	-	-	-	-
13 x	-	-	-	-	-	-	-	-	X	-	-	-

III b - Script 15 Mathematics test score: 31

Problem No.	Category 1 Reading i	ii	iii	Category 2 Comprehension i	ii	iii	Category 3 Ident. of operation	Category 4 Trans-forming	Category 5 Process skills i	ii	iii	iv
1 +	P	P	R	R	-	-	-	R	-	-	-	-
2 -	P	-	-	-	-	-	-	-	-	-	-	-
3 +	P	-	-	R	-	R	-	-	-	-	-	-
4 +	-	-	-	P	-	-	-	-	R	-	-	-
5 -	P	-	-	-	-	-	Adds on	-	-	-	-	-
6 -	-	-	-	-	-	-	Adds on	-	-	-	-	P
7 -	P	P	-	-	R	-	Adds on	-	-	-	-	-
8 -	P	-	-	-	-	-	Adds on	-	-	-	-	-
9 -+	-	P	-	-	-	-	-	-	R	-	-	-
10 x	-	-	-	-	-	-	-	-	R	-	-	-
11 x	P	P	-	-	-	-	-	-	R	-	-	-
12 x	-	-	X	X	-	X	X	X	-	-	X	-
13 x	P	-	-	-	-	P	-	*	R	-	-	-

Key:
- \- no difficulty
- P one prompt required for child to move on in correct direction
- R repeated prompts required for child to move on in correct direction
- X uncorrected error/no response
- * sum transformed by researcher

III c - Script 23 Mathematics test score: 31

Problem No.	Category 1 Reading			Category 2 Comprehension			Category 3 Ident. of operation	Category 4 Trans-forming	Category 5 Process skills			
	i	ii	iii	i	ii	iii			i	ii	iii	iv
1 +	-	-	-	-	-	-	-	-	-	-	-	-
2 -	-	-	-	-	-	-	-	-	-	-	-	-
3 +	-	-	-	-	-	-	-	-	-	-	-	-
4 +	-	-	-	-	-	-	-	-	-	-	-	-
5 -	-	-	-	-	-	-	-	-	-	-	-	-
6 -	-	P	-	-	-	-	-	-	-	-	-	-
7 -	-	-	-	-	R	-	-	-	-	-	-	P
8 -	-	-	-	-	-	-	-	-	-	-	-	-
9 -+	-	-	-	-	-	-	-	-	-	-	-	-
10 x	-	-	-	-	-	-	-	-	-	-	-	-
11 x	-	-	-	-	-	-	-	-	-	-	-	-
12 x	-	-	-	-	R	-	-	-	-	-	-	-
13 x	-	-	-	-	-	-	-	-	-	-	-	-

III d - Script 26 Mathematics test score: 25

Problem No.	Cat. 1 i	Cat. 1 ii	Cat. 1 iii	Cat. 2 i	Cat. 2 ii	Cat. 2 iii	Cat. 3	Cat. 4	Cat. 5 i	Cat. 5 ii	Cat. 5 iii	Cat. 5 iv
1 +	-	-	X	X	-	-	-	-	-	-	-	-
2 -	-	-	-	-	-	-	-	-	R	-	-	-
3 +	-	-	-	P	-	P	P	-	-	-	-	-
4 +	-	-	-	-	-	-	-	-	-	-	-	-
5 -	-	-	-	P	-	-	-	P	R	-	-	-
6 -	-	-	-	R	-	R	R	*	-	-	-	-
7 -	-	-	-	-	-	-	-	-	R	-	-	-
8 -	-	-	-	-	-	-	-	-	-	-	-	-
9 -+	-	-	-	R	-	R	-	-	R	-	-	-
10 x	-	-	-	R	-	-	-	-	-	-	-	-
11 x	-	-	-	R	-	P	-	-	-	-	-	-
12 x	-	P	P	-	R	R	P Adds	-	-	-	-	-
13 x	-	-	-	-	-	-	-	-	-	-	-	-

Key:
- \- no difficulty
- P one prompt required for child to move on in correct direction
- R repeated prompts required for child to move on in correct direction
- X error uncorrected/no response
- * sum transformed by researcher

III e - Scripts 33 Mathematics test score: 22

Problem No.	Category 1 Reading i	ii	iii	Category 2 Comprehension i	ii	iii	Category 3 Ident. of operation	Category 4 Trans-forming	Category 5 Process skills i	ii	iii	iv
1 +	-	-	R	R	-	-	R	R	R	-	-	-
2 -	-	-	-	X	-	-	R	-	-	-	-	-
3 +	-	-	-	-	-	R	-	-	R	-	-	-
4 +	-	-	-	-	-	-	-	-	R	-	-	-
5 -	P	-	-	R	-	-	R	-	R	-	-	-
6 -	-	-	-	-	-	P	-	R	R	-	-	-
7 -	P	-	-	R	-	-	R	-	R	-	-	-
8 -	P	-	-	R	-	-	P	P	R	-	-	-
9 -+	P	-	-	R	-	R	-	-	P	-	-	-
10 x	P	-	-	R	P	-	- Adds	-	-	-	-	-
11 x	-	-	-	R	-	R	R	*	X	-	-	-
12 x	P	P	P	-	R	R	- Adds	-	-	-	-	-
13 x	-	P	P	P	-	-	- Adds	*	-	-	-	-

Key:
- \- no difficulty
- P one prompt required for child to move on in correct direction
- R repeated prompts required for child to move on in correct direction
- X error uncorrected/no response
- * sum transformed by researcher

One profile has been included from each school (see Appendix C for actual transcript and classification of errors). These are not representative as typical of the respective classes but were picked to demonstrate the variety of profiles that exist within a supposedly homogenous group. Analysis of these individual pupil responses did not indicate that the school/class was the only locus of variation although it is likely to be an important one. Each class tended to contain a fairly wide range of response types, with some showing competence in both the procedural and conceptual sphere, some stronger in the procedural, some in the the conceptual, and some strong in neither sphere. It also shows that this type of analysis of a child's ability to solve a problem can provide a rich source for examining a particular child's mathematical understanding. It would therefore be useful to discuss each script individually and comment on: (i) the extent to which each coincides with the general trends and (ii) what can be gleaned from the profile about that particular child's mathematical understanding of the word problems used. A final section will compare the five profiles and relate the mathematics test score to the profiles given.

Script 7. This was one of the first interviews, and was thus carried out without any prompts from the interviewer. This method was changed (see Section 5.2.6). It can be seen that if the child had had difficulty with the comprehension of the problem there would have been no way of ascertaining if the child was in fact capable of executing the sum that was required.

(i) General trends: This child has little difficulty in comprehending the problems but is poor in executing all but the addition problems. She conforms to the general trend in that she has difficulties with her subtraction algorithm and in that she does not understand problem 12.

(ii) Individual understanding of the word problems: Her main difficulty in subtraction is with carrying and she displays a number of Van Lehn's bugs. Interestingly she does not display the same bug every time, even within the same problem. In one problem she incorrectly makes the following error: 0-N=0 in the most right-hand column, and this changes to 0-N=N in the adjacent column. Again within multiplication it is when carrying is required that she makes mistakes. However, she does know how to set up the standard algorithm for either subtraction or multiplication.

Script 15.

(i) General trends: The understanding of problem 12 caused considerable confusion to this child. Despite repeated explanations he failed to understand what was expected and in the end this problem was abandoned without a solution. Problem 3 also caused difficulties in the area of comprehension and this ties in with the general trend. He differs from the general trend in being able to comprehend and solve the more difficult subtraction problems.

(ii) Individual understanding of the word problems: This child has a great problem with reading the problems and required much help. As can be seen comprehension is on the whole not problematic once the reading difficulty has been sorted out. He does, however, use his own idiosyncratic methods for solving the problems as the entries to Category 3 (Identification of Operation) show. When subtraction is intended he adds on. He is quite lost when required to carry out a subtraction or multiplication using the standard algorithms: he does not know how to transform these type of problems into the standard form. The entry for problem 13 shows this. For all the other problems he avoids doing this by carrying out the sum in his head and not using paper and pencil. Problem 13, however, proved too complicated for mental arithmetic and his idiosyncratic methods failed him.

Script 23.

(i) General trend: This shows the transcript of a child who was very competent and it thus deviates from the general trend.

(ii) Individual understanding of the word problems: This child shows a thorough understanding of the word problems. The two entries in the comprehension column refers to lack of understanding of two words within the problem. However, this did not prevent her from reaching a correct solution.

Script 26.

(i) General trend: This child conforms to the general trend in that he finds problems 5, 6, 9 and 12 problematic.

(ii) Individual understanding of the word problems: comprehension is on the whole more problematic for this child than are process skills. Had this child not been helped to comprehend the problems he would have no chance of showing that he in fact is reasonably competent with the procedures. Subtraction does, however, cause some difficulties both in the area of process and transformation.

Script 33.

(i) General trend: This child has greater difficulties than the majority of the children in the population.

(ii) Individual understanding of the word problems: This child has difficulties right across all categories. She needed much help not only to read and understand the problems but also to transform and carry out the calculations. Her understanding of the standard multiplication algorithm was poor, however, she was able to represent and calculate the multiplication problems as repeated addition. This was after it had been suggested to her that she used a diagram representing the quantities involved. (For example, for 4 classes with 32 pupils in each, a classroom was drawn for each class with the number of pupils entered into each classroom drawn).

Comparison of scripts. (For ease of reading each child will be referred to by the number of his/her script.) As can be seen, there is quite a bit of variety. Child 23 undoubtedly stands out as very competent and she was indeed in the top maths groups in her class. The other four scripts show that there are a number of differences between the pupils in their understanding of mathematics. 15 is a child who seems to comprehend the problems but is very poor in the standard algorithms. He can nevertheless achieve a correct solution if the numbers involved are simple enough through using his own idiosyncratic method of mental arithmetic. He is also possibly hampered when working on his own by his reading difficulty. He was in group 3 in his class. Child 26 on the other hand had fewer reading and process difficulties but more in the area of comprehension. Where his lack of process skills failed him he did not have any idiosyncratic methods to fall back on. Child 33 could perhaps be termed as generally a slow learner in that she experienced difficulties in all the areas required for a solution.

As will be seen from the results, many of the pupils involved in this research experienced difficulties when trying to solve the thirteen word problems. Two points emerge: firstly that for many there was little apparent retention of the recently learnt concepts. These problems were not new to the children; they had encountered most of them in class during the two months preceding this research. Some of the children had just "learnt" how to deal with some of the problems the week prior to the interview. The effectiveness of this learning could be questioned by considering the following incident: one child remarked on being asked to do a particular problem "I did that one last week - but I didn't get that answer". The only recollection he had of doing this particular problem was that the answer he got in class did not match the one he got when doing it with the researcher. How he had got the answer in class totally evaded him and he was not able to work it out from reading the word problem. There was only one child in the whole sample who managed to work out all the problems correctly without any help (see script 23).

Secondly, the majority of these children were not considered by their teachers to have learning problems. So, why do they find these word problems so difficult? To attempt to pursue a single, dominant factor that affects the difficulty of a word problem would be simplistic. A problem considered difficult by one child does not necessarily cause difficulties for another child. The factors affecting the difficulty may be located within the word problem - and here it may be language, mathematical processes, graphical or layout details that are causing the difficulty; within the child; or within the teaching methods, or within a combination of any of these. A number of these factors have been discussed above. The second phase of the main study will look more closely at the factors that affected the difficulty of some of the word problems used for this research. This will be discussed in the section on "Further research". Having looked at the actual results the issues that arose out of this phase of the research will now be explored.

5.4 Issues Arising.

The points to be looked at in this section are:

(i) the concept of the "average pupil"

(ii) the SPMG teachers' handbook

(iii) problem representation

(iv) sex differences

(v) the use of concrete materials

(i) The "average" pupil. It has already been suggested when discussing Table III that although the children in this study fell within the "average" there seemed to be considerable differences in the type of difficulties they were experiencing. It would be useful first to look at a framework for analysis of the children's responses and then to look at differences and similarities in these responses using the suggested framework. The implications for the teaching offered these pupils can then be discussed.

Framework for analysis and its links to this research. It would be useful here to use the framework provided by Hiebert & Lefevre (op.cit.) and other researchers for analysing mathematical competency. They distinguish between conceptual and procedural competencies. Conceptual competency refers to the interconnected web of knowledge that is generally thought to represent understanding of a problem whilst procedural competency refers to the skills required to execute the problem correctly. It can be suggested that these two competencies are reflected in the error analysis used in this research: Comprehension and Identification of Operation corresponding to conceptual competency; Transforming and Process skills linking to the procedural competency. Both competencies are generally essential for a correct solution to a problem though it is sometimes possible to solve the problem without understanding its underlying structure by using a memorised routine. This would show procedural competency without conceptual competency. Conversely a child may understand what to do but not be able to carry it out. Here conceptual competence is in evidence but procedural is lacking. The question then turns to what similarities and differences do these "average" pupils display in these different competencies.

Differences and similarities. Looking at the five scripts in Table III it is suggested that one pupil is effective in both type of competencies (script 23); one seems to be weak in both areas (script 33); one seems particularly weak in procedural competence (script 7); one is weak in procedural competence, particularly in relation to the standard subtraction algorithm; the final one is weak mainly in conceptual competence, but also in procedural competence when required to subtract. This evidence suggests four different groups:

1. competent in both areas (script 23)
2. competent in procedural (script 26)
3. competent in conceptual (script 15, 7)
4. incompetent in both areas (script 33)

Looking at further evidence it is interesting to note that the two pupils with script 23 and script 15 had the same score on the mathematics test and yet showed quite different profiles for understanding of the problems that formed part of this study. Therefore there were differences within the sample but similarities also exist in that amongst the forty children interviewed a large proportion found certain problems very difficult. There are also similarities to be found, as already suggested, between a number of these children from different classes in that they will tend towards one of the above groups and thus be similar to some of the children within this population.

Implications for teaching. This research is of a very limited nature and hence can only make tentative suggestions in this area. However, it does seem fair to suggest that current practice of "ability" groupings within the classroom may not in fact reflect the competencies of the pupils effectively. As an example, child 15 was in the bottom group in his class. Yet he understood two of the most difficult problems and could solve them. He was, however, unable to use the standard subtraction and multiplication algorithms. This child would possibly benefit from learning to be more effective procedurally to allow his competence in this area to catch up with his conceptual understanding. The same could perhaps be suggested for child 7, whilst child 26 would benefit from learning that would increase his conceptual understanding of word problems. It has been mentioned earlier that the pupils from school 1 seemed to have the firmest grasp of the recently learnt material. This class was not organised in ability groupings in mathematics. Thus perhaps another suggestion that ability groupings are not essential for effective teaching.

(ii) SPMG handbook. The handbook is organised to give teachers advice on how to teach specific pages and in some instances specific topics. One such topic is the "subtraction using the decomposition method". This section will consist of three subsections:

a) the advice given by the handbook

b) the type of advice that could be offered (based on the observed difficulties)

c) summing up

a) The advice given by the handbook. In general, drawing from the pages that relate to the word problems used, the handbook states that word problems "are very important" and it stresses that the children "should be given extra problem-type examples". In another place it states that a particular page requires "good reading skills".

On the topic of subtraction it states that the children should be taught to use the subtraction technique requiring decomposition. Earlier pages covering the topic of subtraction emphasises the language of subtraction - with the stated intention that the child learns to recognise the words used and learn to subtract when s/he sees it. There seems to be an inherent danger in this suggestion which will be discussed below. On the whole the advice given focuses on the arithmetic operation required and how to teach this.

b) The type of advice that could be offered.

First the advice relating to reading the problems seems quite vague. Given that the reading of mathematical language differs from the reading of ordinary English (see, for example, Kane, 1970, Shuard & Rothery, 1984) in some very specific ways, perhaps more precise guidance could be given. There is a very low level of redundancy in mathematical language so every word needs to be read carefully. There was evidence of children not understanding the problem statement because they did not read every single word (see, for example, problem 12). A word such as "each" being left out can alter the whole meaning of the problem.

There are also a number of words and phrases that have different meanings in mathematical English. One such example is the word "difference" or phrase "difference between". When a child is being asked to find "the difference between" s/he is being asked to compare two quantities and state by how much one differs from the other. Lack of this type of understanding was in evidence (see problem 5). A number of children responded with "one is higher/ bigger" when they were asked to "find the difference". As shown above, the handbook seems to encourage that the child learns to respond with a particular mathematical operation when s/he encounters this word, or that phrase. This seems to encourage the idea of the children responding to verbal cues and thus possibly bypassing the analytic reading of the problem. Evidence of this behaviour was found with a number of children who responded with "because the teacher told me" when asked why s/he had carried out a subtraction when reading "difference between". A number of these children had then added when confronted with "more than". In other words, they lacked conceptual understanding of the language of subtraction, responding instead to the stimulus of word recognition.

Problem 12 showed that a number of children have difficulties with ordinary English words such as "clumps", and problem 1 showed that lack of contextual knowledge caused some confusion. No mention is made of any of these type of difficulties possibly arising.

c) **Summing up.** It thus seems that little specific advice is given in relation to the skills required to solve word problems. The advice given tends to concentrate on arithmetic operation. Little is said about language that emphasises the differences between mathematical and ordinary English. For the inexperienced teacher the advice must seem rather vague, for the experienced, competent teacher probably not particularly helpful.

(iii) Problem representation and recording. This relates to evidence from Table III in particular. It has been noted that a number of pupils used their own idiosyncratic methods to solve the problems and did not use those that were intended by the problem writer. The fact that a number of children were able to understand the problem and solve them effectively without the standard algorithm suggests an underlying, informal understanding that is very useful. However, as already noted these methods were not helpful when the arithmetic became too complicated. It seems to be essential that teachers are aware that children use these methods and ensure that they link these to the standard algorithms to provide the child with a thorough understanding of both. The child who can effectively use mental arithmetic has a powerful tool when it comes to estimating answers. This can be very useful, for example, when using calculators, to ensure that no unrealistic mistakes are made. A correct answer that is assumed to signify understanding of the standard algorithm and is recorded in the child's record (in the case of SPMG in the space provided on the front of the workbook) may give an inappropriate picture of the child's mathematical understanding, if the child uses an idiosyncratic and informal method to reach the solution. If this record is then handed to another teacher. this teacher may mistakenly believe that the child is reasonably well versed in using the standard algorithms in question.

It is interesting to note here that, during the research period, two out of the five classes had a change of teachers during the year and one class had several changes. Thus it is essential that the teachers are aware of these idiosyncratic methods of the pupils and that they ensure that the standard algorithms are well practised.

(iv) Sex differences. The exploratory study found sex differences in the type of difficulties that were experienced by boys and girls. It suggested that girls tended to have greater difficulties in the areas of comprehension whilst boys had greater problems with process skills. No such differences were found in Phase 1 of the main study, so this is not considered a topic worth pursuing.

(v) The use of concrete materials. It was suggested in the exploratory study that the use of Dienes material might have a beneficial effect on the children's ability to solve the word problems. SPMG encourages the use of concrete materials and considers it a useful link between the practical world and the formal symbolic representation of mathematics. The textbook contains pictorial representations of concrete materials and encourages the children to use concrete materials as shown in the picture, as an aid to computation. It was found, in this study, that the materials were used by the children in two different ways: firstly by using single units (and if necessary single tens) each to represent one. The child may be wishing to find out 9 from 17 (e.g. as part of the calculation 87-39). S/he would count out 17 units (using tens if there were not enough units available) and take 9 away and thus find the answer. In this way the material seemed to provide the child with an effective aid to a correct calculation. Secondly, the material was used to represent subtraction sums with the various parts (hundreds, tens and units) of the material used as intended to represent the numbers. However, when using the material this way the children invariably failed to reach a successful solution. They lost sight of the goal whilst manipulating the material and failed to keep the different aspects of the problem apart. It is interesting to note here Van Lehn's (1986) bug "Borrow-Unit-Difference". This bug represents the following error: the pupil works out what is needed for the top digit to be equal to the bottom digit and then decrements the adjacent top digit to the left by that amount. This bug, he suggests, is mainly found amongst children who have used concrete materials. He suggests it occurs due to a lack of adequate links between the concrete and symbolic representations.

The extent to which the manipulation of material was encouraged within the classroom was not ascertained, though all the children had access to this type of material. The extent to which the depicting of concrete materials in the textbook is helpful is questionable. The children who did not need the material could probably follow these pictures. The children who were not able to use the material effectively would probably find it very confusing to try to read their way through an imagined manipulation of materials. In fact, it is debatable whether something intended as a practical experience has any place in a textbook where it can become a pretended rather than actual practical experience. If it is intended to show the teacher how to use the material it is probably better left in the handbook. The extent to which those children who find mathematics difficult anyway are helped by the use of a single object used to represent 100 when another single object is used to stand for 1 is maybe questionable. If children are to use concrete materials effectively in subtraction sums they need careful teaching. Research by Resnick, quoted in Van Lehn (1986) suggests that children are unlikely to make connections between the concrete representation and the symbolic representation unless this link is specifically drawn to their attention. This could be seen as a stance against the importance of practical experience in learning mathematics. It should, however, not be interpreted as such. The points to be questioned are: if certain aspects of a material created by mathematically skilled adults actually do provide children with the kind of help that is needed; and if the materials are being employed in the classroom in the intended manner. The use of concrete materials was seen as helpful in alleviating memory load when the material was used to represent single units but not when it was used to calculate sums where the material represented larger numbers and these representations were linked to the intended values of the materials.

5.5 Further research - Problem specific difficulties.

Previous research that has looked at difficulties in word problems will first be looked at in relation to Phase 1 of this research. The next stage of this research, Phase 2, which explores further the difficulties of some of the problems used in Phase 1 can then be discussed.

Other research in this area and its relationship to this research. The idea that particular, definable aspects of a word problem are responsible for its difficulties has been studied in great depth by amongst others Nesher & Teubal (op.cit.), Jerman (op.cit.).). Others, such as Kane (op.cit.) and to some extent Shuard & Rothery (op.cit.) have suggested that the difficulty lies within the translation and presentation of the problem. Those researchers, known as the structuralists, looking for clues within the problem, have suggested such factors as the problem length, absence or presence of verbal cues and the number of different arithmetic operations required for the solution as affecting problem difficulty. Kane emphasised the effect of language and suggested that a lack of awareness of the differences between mathematical and ordinary English may be a contributory factor to the difficulties of word problems. Most of the structuralist research has used researcher created word problems that intend to offer effective control of all the variables except those under study. This research has looked at the word problems that the children actually encounter in the classroom and thus has no such controls either on problem length or the type of words used. However, it could be suggested that when attempts are made to control, for example, problem length one may introduce into the language an artificiality which might affect the children's interpretation of the problem statement. The intention of the textbook writers is that word problems offer an opportunity to practice mathematics within a realistic setting (see also the discussion in Chapter 1, p. 6). An example of this researcher creation of word problems to test for effect of extraneous information is "A tailor sewed out of a 56 m long piece of material 7 identical suits with a modern and handsome cut. Find out how long was the material required for each suit". This problem was aimed at the 13 - 15 year age group. It does not seem that this type of problem offers a particularly realistic way of practising mathematics. The younger the children involved in the research the more important it becomes to be aware of the limited knowledge of language that the children possess (see e.g. Donaldson, op.cit.). The SPMG textbook recognises these language limitations to some extent in that it uses very simple sentences.

Phase 2. Following Phase 1 of the main study it was decided to look at specific aspects of the problems as one possible cause of difficulties. In the light of the discussion in the previous section it was decided to rewrite the problems using the difficulties identified in Phase 1 as the basis for the rewritten material. This was in preference to drawing on structures identified by other researchers as affecting problem difficulty. It was suggested that these problems created by other researchers may not actually relate well to the type of problems the children meet in their textbooks. It is intended that the new problems thus created would be more closely related to the type of problems that the children encounter during their normal mathematics lessons. The second phase of the main study then consists of selecting a smaller number of problems for further study with the same sample group. The selection of these problems and the effects noted will be discussed in the next chapter.

5.6 Conclusion.

It has been found that many of the difficulties experienced by the children in the exploratory study were also in evidence in the main study, where a more homogenous population formed the sample. A number of issues arising from Phase 1 of the main study have been discussed: the error analysis has been further refined and the interview procedure slightly changed, the concept of the "average" pupil and the possible effects on problem solving of different teaching approaches have been discussed. The usefulness of the handbook, the acuity of the progress recording within the SPMG scheme as well as the effect of concrete materials were examined. It has been noted that sex differences apparent in the exploratory study were not in evidence in this part of the main study. In order to try find out more about specific aspects of difficulties that may be located within the problems and their presentation it was decided to carry out a further study. This formed Phase II and involved the rewriting of a number of problems. These were then given to the same population to see what effects there were, if any. The selection, rewriting and presentation of these problems forms the topic for the next chapter.

CHAPTER 6 – MAIN STUDY: PHASE 2

6.1 Introduction.

Phase 1, in the last chapter, used the task based interview to look at the difficulties experienced by pupils when trying to solve SPMG word problems. To try to pinpoint more precisely the aspects of the actual problems that were causing difficulties a further study was conducted. Six of the original problems were rewritten and presented to the same population. The rationale behind the rewritten problems, their presentation to the children and the results of these presentations form the contents of this chapter.

As the population used was the same as in the previous chapter its details need not be repeated here, except to explain minor changes. The main differences between Phase 1 and Phase 2 are in the way the data was gathered and the problems were used, and these two aspects will be discussed in detail. The chapter thus takes the following form: an initial section on method will

(i) give the overall design of this phase of the study,

(ii) describe the choice and rationale behind the rewritten problems,

(iii) show the organisation of these rewritten problems into three test papers

(iv) indicate the materials used,

(v) comment on the slight changes in the population,

(vi) explain the procedure employed.

The results and discussion will examine:

(i) the responses to the original problems compared to those of the rewritten versions. Chi-squared tests were used to determine any statistically significant differences.

(ii) the general trend of the results of the chi-squared tests and the type of factors that seem to create the greatest difficulty.

(iii) the ability of verbal cue theory and schema theory linked to the notion of a cognitive workbench, to provide an explanation for the children's responses.

(iv) differences and similarities between the schools

(v) differences between Phase 1 and Phase II in the method of data gathering.

The conclusion sums up and suggests further areas to be explored.

6.2 Method.

6.2.1 Design. Phase 1 of the main study looked at difficulties experienced by children in the middle ability range. This study sets out to look in further detail at six of the word problems used in Phase 1. In an attempt to locate more specifically the possible loci of difficulties within these six problems, a number of rewritten versions were created. These rewritten problems were divided into three different tests which were presented to the children, one test at a time, once a week for a period of three weeks. The analysis of the results looked at correct/incorrect responses. The incorrect responses were further examined to provide a division into conceptual (comprehension) and procedural (transforming and process skill) difficulties. This allowed for a consideration of how the type of errors related to the nature and structure of the problem.

6.2.2 Rewritten problems. It was decided to select six problems for this phase. Each problem was to have no more than four different word problem versions and some would have less. Any more problems than this might create difficulties in the testing stage as the children tend to get bored with too many similar problems. This was in fact the case towards the end of the data gathering and a few children had to be encouraged to complete the task. (However, they were no more bored with this than they were with some of their ordinary mathematics tasks, and the children that had to be cajoled were the ones that the respective teachers needed repeatedly to encourage to work.)

The selection of problems to be used was governed by the difficulties the children had experienced during Phase 1, with one difference: the possibility of "cognitive overload" was introduced and will be further explained below. Those problems that had created the greatest number of difficulties in Phase 1 were used with the exception of one: Problem 1 was among the six most difficult, but was considered unrepresentative of the type of word problems that children of this age group generally meet within school mathematics. This problem required the addition of a string of digits, involving a heavy demand on working memory skills. The main error was one of miscalculating slightly or forgetting to add the last digit on. Usually addition word problems for this age range involve the addition of two numbers using column addition.

Rationale behind changes made to the original problems. The problems were rewritten to explore further the difficulties experienced by children when trying to solve the original problems. The difficulties varied but the following categories of difficulties shown below emerged as the most prominent during the analysis of the Phase 1 data. These were used as a basis for creating the rewritten problems. The first category of problem relates to conceptual or comprehension difficulties and the second and third categories, of transforming and process skills, relate to procedural difficulties. The terms comprehension, transforming and process skills have been retained as they form a link with the analysis in Chapter 5.

Comprehension difficulties.	(i)	lack of contextual knowledge
	(ii)	reliance on keyword/verbal cue
	(iii)	lack of understanding of mathematical usage of particular word
	(iv)	confusion over graphics
	(v)	inability to identify information relevant to the problem statement and necessary for the solution of the

	problem - in the problems this was mainly due to layout of problem or number of arithmetic operations required for solution
Transforming difficulty.	inability to transform a problem statement into a suitable mathematical algorithm to enable solution
Process skills difficulties.	inability to calculate solution using the standard algorithm, here mainly due to inability to carry

Cognitive overload. An additional category was considered. This was suggested by the fact that a number of children found the problems easier to understand if the numbers involved were much reduced but the problem statement remained the same. This category has been termed "cognitive overload". The suggestion is that if both comprehension and process aspects of the problem are difficult the total effect is akin to a "gestalt" effect; that is, the total amount of difficulty would be greater than the sum of the two individual parts that make up the problem. This idea of cognitive overload links into Britton, Glynn & Smith's (1985) idea of a "cognitive workbench". This is discussed further in Section 6.3.4.

These categories of difficulties formed the basis for the rewritten problems. In each of the rewritten problems it was intended to isolate, and present in a prominent way, one of these difficulties, so that their relative effects might be considered. As already mentioned, the categories derived from the type of difficulties experienced for that problem by the children in Phase 1.

The six original problems are shown below and the target difficulties indicated. The rewritten problems follow each original problem. The numbering of the original problems has been retained so that reference can be made to the results in Chapter 5. The rewritten problems are numbered 1 to 22, preceded by an R to indicate that it is a rewritten version. The parts of the problems which have been changed are indicated by italics. A number of other alterations have been made to names and numbers to avoid as far as possible the effect of practice and to make the problem seem "different" to the children. These alterations are not shown in italics.

Problem 2: Jim enters the 80 metres race and is given a start of 13 metres. How far does he have to run?

Difficulties in this problem were mainly in the area of comprehension: (i) lack of contextual knowledge, (ii) reliance on keyword; and with process skills.

Rewritten problem R1 deals with contextual knowledge by inserting an explanatory phrase:

R1. Jim enters the 80 metres race and is given a start of

13 metres. *This means he does not have to run all the 80 metres.* How far does he have to run?

R2 substitutes the keyword "given" with one less suggestive of addition. Other research (e.g. Nesher & Teubal, op.cit.) indicates that "given" is a word likely to suggest addition:

R2. Alison enters the 90 metres race and *starts 12 metres in front of the others.* How far does she have to run?

R3 deals with process skill difficulties by slightly altering the numbers so that no carrying is required:

R3. David enters the *85 metres* race and is given a start of *13 metres.* How far does he have to run?

Problem 3: One morning 37 boys and 46 girls go to the library.
That afternoon 39 boys and 59 girls go.

How many a) boys b) girls go to the library that day?

Difficulties in this problem were mainly in comprehension - (v) identifying relevant information. The layout with (a) and (b) interwoven with the text seemed to encourage some children to add up all the numbers. Further evidence that this layout was confusing is shown in the transcripts. A number of children, who chose to read the problems aloud omitted the (a) and (b) and inserted an "and" between boys and girls. Only one change was made to reflect this difficulty:

R4. One morning 37 boys and 46 girls go to the library. That afternoon 39 boys and 59 girls go.

a) How many boys *go to the library that day?*

b) How many girls go to the library that day?

[as only (a) was used in Phase 1, the children were asked to complete only (a)]

Problems 5 and 6:

As these two problems were the (a) and (b) subsection of the same problem they have been retained as such. Problem 5 is (a) and problem 6 is (b).

During a game of darts Billy King had scored 187 and Jock Scott 223.

a) What is the difference between these scores?

b) How many more does Billy need to make 301?

These problems created many difficulties: in the area of comprehension: (ii) reliance on keyword, (iii) lack of understanding of mathematical use of word; in the area of transforming; and that of process skills. As these were two problems that posed a lot of difficulties the "cognitive overload" category of difficulties was tested on these problems. R5 - R9 relate to problem 5, and R10 - R14 form rewritten versions of problem 6. R5/6 and R10/11 look at the comprehension aspects of the problems:

R5/10 During a game of darts Bob Smith had scored 187 and David Brown 223.

R5 *How much less does* Bob Smith *have than* David Brown?

R10 *What is the difference between* Bob's *score and a score of 301?*

R6/11 During a game of darts Colin White had scored 167 and Neil Stewart 213.

R6 *How much more does* Neil Stewart *have than* Colin White?

R11 Colin *wants to make 303. How many less than 303 does he have?*

R7 and R12 deal with transforming by presenting the child with sums transformed in the manner intended by the particular problem to which it is related:

R7. *263*
-197

R12. *402*
-237

R8/13 looks at process skills, and removes the need to carry:

R8/13. During a game of darts Chris Smith had scored *123* and Bill Brown *235*.

R8 What is the difference between these scores?

R13 How many more does Chris need to make *255*.

and R9/14 cognitive overload:

R9/14. During a game of darts Mike Wood had scored *36* and Jack Macdonald *43*.

R9 What is the difference between these scores?

R14 How many more does Mike need to make *59*?

Problem 9: Farmer Till had 210 sheep. At the market he sold 88 and bought 25. How many sheep has he now?

This problem evoked difficulties mainly in comprehension: (v) identifying relevant information; and in the area of transforming and process skills. As the results from Phase 1 suggested that it was a complex problem it was also examined from the point of cognitive overload.

R15 looks at comprehension:

R15. a) Farmer Till had 210 sheep. At the market he sold 88 sheep. *How many sheep has he now?*

b) He then bought 25 sheep. How many sheep has he now?

R16 and R17 examines transformation with a sum related to the types of sums intended by the problem:

R16. *220*

-76

R17. *114*

+33

R18 investigates process skills:

R18. Farmer Brown had *198* cows. At the market he sold *86* and bought *33*. How many cows has he now?

and R19 cognitive overload.

R19. Farmer Macdonald had *60* sheep. At the market he sold *5* and bought *3*. How many sheep has he now?

Problem 12: In a garden there are

2 clumps each with 145 snowdrops,

3 beds each with 72 tulips, and

5 beds each with 50 daffodils.

a) How many snowdrops are there?

b) How many tulips are there?

c) How many daffodils are there?

Difficulties in this problem were mainly in the area of comprehension: (iv) the use of graphics, and (v) identifying relevant information; and in layout.

R20 deals with the inessential graphics by removing them:

R20. In a garden there are

2 clumps each with 145 primroses,

3 beds each with 72 lupins, and

5 beds each with 50 poppies.

a) How many primroses are there?

b) How many lupins are there?

c) How many poppies are there?

R21 simplifies the problem by reducing the information given:

R21. In a garden there are

2 clumps each with 135 snowdrops.

How many snowdrops are there?

and R22 alters the layout:

R22. In a garden there are 2 clumps each with 125 snowdrops, 4 beds each with 69 pansies, and 6 beds each with 73 roses.

a) How many snowdrops are there?

b) How many pansies are there?

c) How many roses are there?

[as only (a) was used in Phase 1 the children were only asked to complete (a)]

6.2.3 Organisation of the rewritten problems.

Three tests were created from these problems to allow presentation of one test each week (see App. D). The tests were lettered A, B, and C. Each test contained only one rewritten version from each problem, so that similar problems were not presented together but had a week's interval between presentations. One exception had to be made as there were four rewritten word problems for problems 5 and 6. Thus two rewritten versions had to be contained within the same test - B. These three different tests were further divided into three different presentations, to mitigate the effect of order as far as possible, and allow for the effects of fatigue or practice:

(i) pure subtraction problems were alternated with those requiring other operations (A1, B1, C1)

(ii) pure subtraction problems formed the first part of the paper (A2. B2, C2)

(iii) pure subtraction problems formed the final part of the paper (A3, B3, C3)

Graphics from the relevant SPMG textbook pages were used to create an effect of similarity to the textbook. For technical reasons only black and white graphics could be employed.

6.2.4 Materials used. Rewritten problems, paper and pencil.

6.2.5 Subjects. There were thirty-seven subjects. All the original children were still at School 1 and 2, but at schools 3, 4 and 5 one child from each class had left and were therefore not able to participate in this phase of the study.

6.2.6 Procedure. The rewritten problems were presented to the children at weekly intervals. All the children were tested at the same time, either seated together at one large table or in two smaller groups, depending on the furniture available. As the children were given different versions of the rewritten material they were not able to copy answers. If a child was absent then s/he joined the group the following week. In case of prolonged absence the child was seen at a later date by the researcher. All the thirty-seven children completed the tasks and no child had less than a week in between each session.

Each child was asked to complete one version each of A, B and C. This means that every child was asked to solve all of the rewritten problems but in different order to minimise the effect of practice or fatigue. The presentation of the different tests and versions was balanced to ensure an almost equal number of presentations. As there were 37 children in the sample totally equal presentation was not possible.

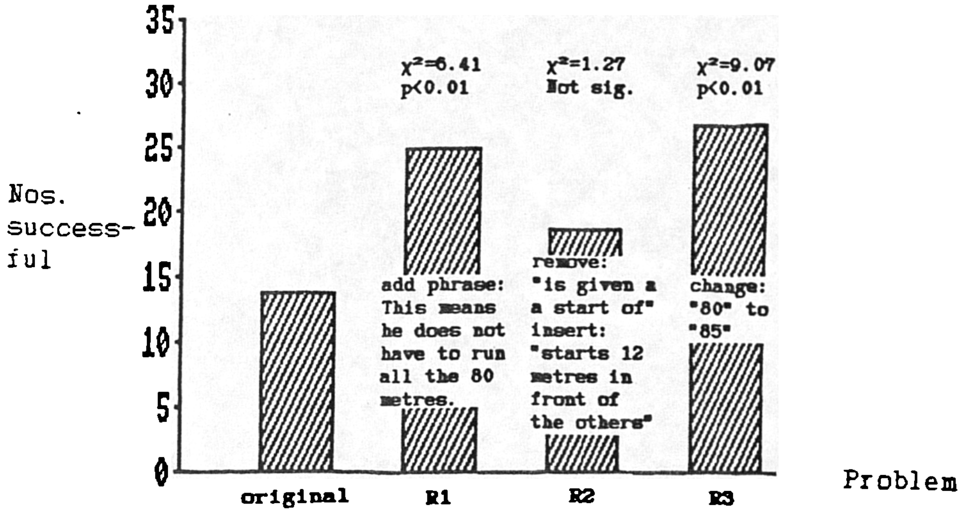
6.3 Results and Discussion.

6.3.1 Chi-squared tests comparing the scores on the original problems with those on the rewritten versions. As the difficulties represented in the rewritten versions were varied, each original problem will be looked at separately in relation to its rewritten versions. For each problem the original problem is restated and the type of difficulties associated with the problems is intimated. As the rewritten versions are discussed at length in Section 6.2.2 only a brief outline of the changes made is indicated. The results of the chi-squared tests are displayed in bar charts and the actual changes made to the problems are shown here. The results then consider the overall picture emerging with differences and similarities highlighted as appropriate.

Problem 2: Jim enters the 80 metres race and is given a start of 13 metres. How far does he have to run?

The first two rewritten problems, R1 and R2, examined conceptual difficulties, whilst R3, the third rewritten problem looked at difficulties relating to procedural skills.

Figure 1a: Chi-squared tests comparing problem 2 with each rewritten version. (DF=1)



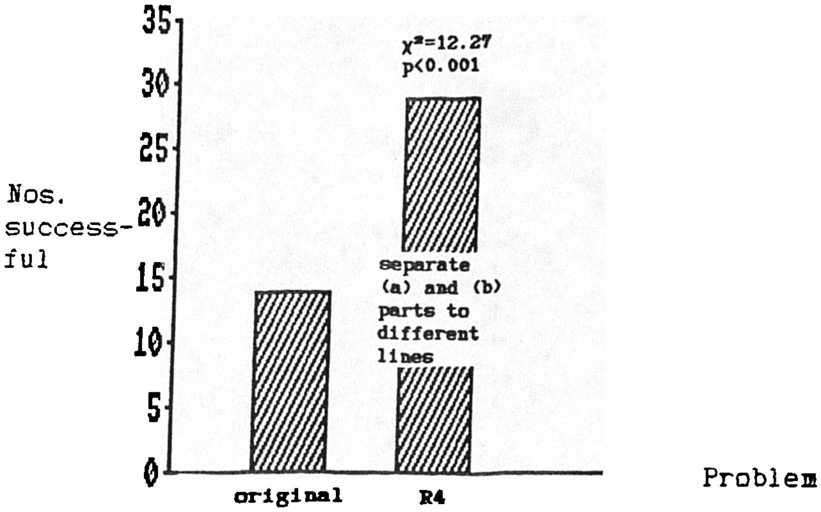
As can be seen from the figure above all the rewritten problems indicate an improvement in the children's performance. However, the improvement elicited by R2, which removed the keyword was not statistically significant, whilst that of R1, inserting an explanatory phrase, was. This suggests that the conceptual difficulty the children experienced for this problem stemmed more from the lack of understanding what "given a start of" meant than from the use of "given" as a keyword. R3 also showed a significant difference from the original ($p < 0.01$). The insertion of an explanatory phrase thus helped the children to achieve a correct solution. However, changing the numbers to remove the need to carry also showed a much improved performance. This suggests at least two variables that affect the difficulty of this problem: one stems from the conceptual and one from the procedural aspect of the problem.

Problem 3: One morning 37 boys and 46 girls go to the library.

That afternoon 39 boys and 59 girls go.

How many a) boys b) girls go to the library that day?

Figure 1b: Chi-squared test comparing problem 3 to its rewritten version. (DF=1)



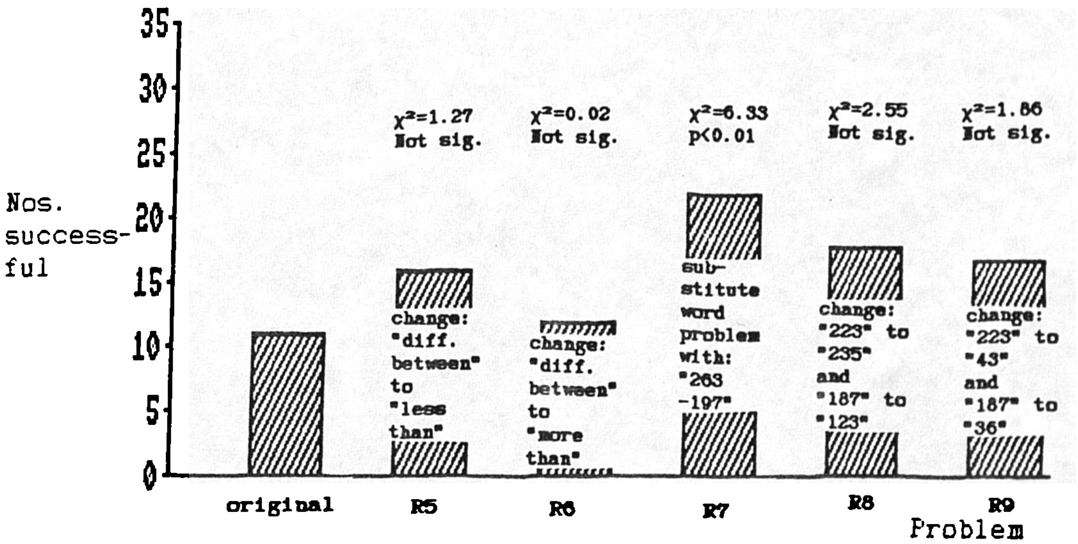
There was only one rewritten version to this problem and it separated the (a) and (b) parts of the problem. The figure above shows a highly significant difference between the original and rewritten version. This result suggests that separating the problem out makes it considerably easier to solve. The idea of a "cognitive workbench" as part of the memory function has been suggested by Britton et al (op.cit.). The idea essentially means that only a limited number of concepts or variables can be handled at any one time. Thus the idea of "overload" (as suggested earlier, in Section 6.3.2) can be brought into operation when the number of variables becomes too great for the workbench to cope. It might be that these results show evidence of this type of overload. This idea will be referred to in relation to other problems and discussed further in Section 6.3.2.

Problem 5: During a game of darts Billy King had scored 187 and Jock Scott 223.

a) What is the difference between these scores?

This problem created a great number of difficulties and there were four rewritten versions in the form of word problems and one with numbers only set out in the standard algorithm. R5 and R6, the first two word problems related to conceptual difficulties. The third word problem, R8 and the numbers only sum R7, looked at procedural skills and transforming. R9 examined the additional category of cognitive overload.

Figure 1c: Chi-squared tests comparing problem 5 with each rewritten version. (DF=1)



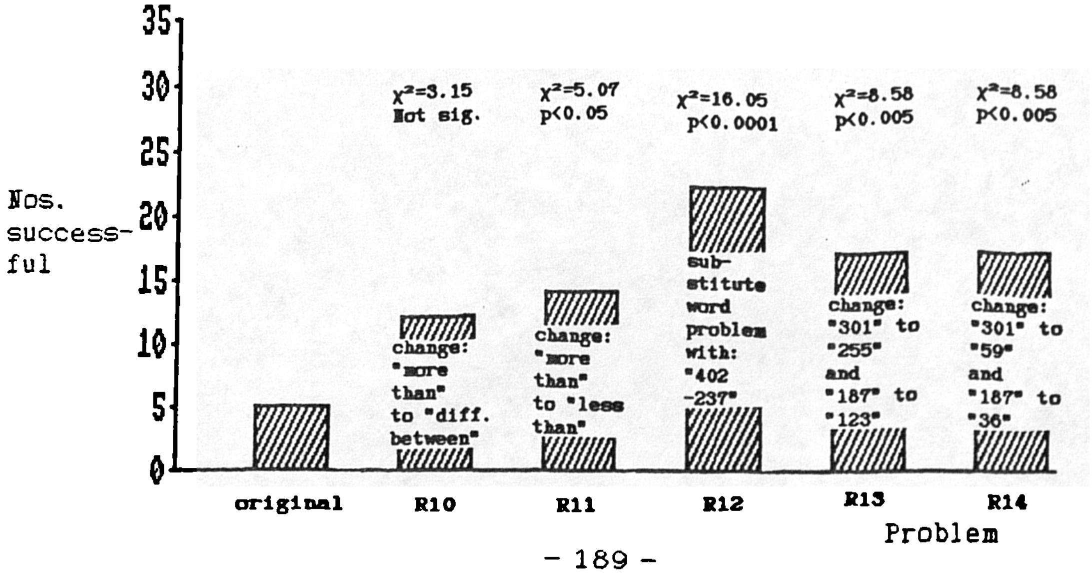
Only one rewritten version was significantly easier than the original problem. This was R7 which used only the numbers from the problem set out in the standard subtraction algorithm. A slight but statistically insignificant improvement in performance was suggested by the results for all of the rewritten versions except for R6. This problem involved the use of the term "more". This, it has been suggested (Nesher & Teubal, op.cit.) acts as a cue for addition. Possibly this was the effect in evidence in this problem. The evidence of support for the verbal cue theory will be discussed below in Section 6.3.3 (i).

Problem 6: During a game of darts Billy King had scored 187 and Jock Scott 223.

b) How many more does Billy need to make 301?

This problem is related to problem 5 and the rewritten versions, like it, used four rewritten word problems and one which looked at the numerical aspect of the problem in isolation. The first two versions, problems R10 and R11 looked at conceptual difficulties. The transforming and procedural aspects of the problem were investigated in a numbers only sum, R12, and in a word problem, R13. Cognitive overload was examined through problem R14.

Figure 1d: Chi-squared tests comparing problem 6 to its rewritten versions. (DF=1)

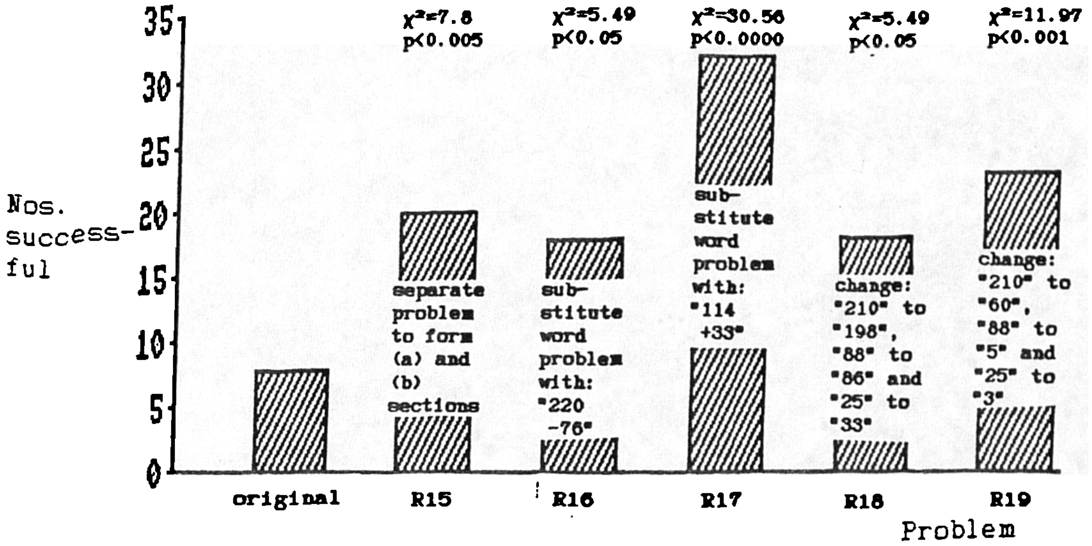


Only one of the rewritten versions did not produce a statistically significant result compared to the original and that was the version using the phrase "difference between" (R10). Changing the figures, removing the need to carry, but retaining the figures in the hundreds (R13) proved as easy as changing the figures to be less than 60 (R14). Thus it can be suggested that handling the larger numbers is not as difficult as having to carry. The rewritten versions relating to conceptual changes produced least improvement. However, the rewritten version that produced the greatest significant change in ability to solve it was the sum without the word problem. Thus the procedural skills are available but when called upon in conjunction with a complex statement using the special language of subtraction difficulties arise. Perhaps further evidence for the notion of a "cognitive workbench". Again an overall improvement in the responses was in evidence.

Problem 9: Farmer Till had 210 sheep. At the market he sold 88 and bought 25. How many sheep has he now?

There were five rewritten versions for this problem. The first one, R15, looked at conceptual difficulties. Transforming difficulties were looked at in R16 and R17. Lack of process skills was investigated by R18, and R19 examined the concept of cognitive overload.

Figure 1e: Chi-squared tests comparing problem 9 with each rewritten version. (DF=1)



Statistically significant results were produced by all the rewritten versions compared to the original. Not surprisingly the improvement elicited by the addition sum proved to be highly significant. Other research has shown these type of sums to be the easiest type of problems. However, the pure subtraction sum proved more difficult than the problem separated into two sections (R15) and the cognitive overload (R19). This suggests the possibility that a number of children reached a solution for the word problems without using the standard algorithm. A word problem may allow for more diversity in strategy than a sum represented in the standard subtraction algorithm. This was suggested by one of the children who responded to 263 with "I can't do that".

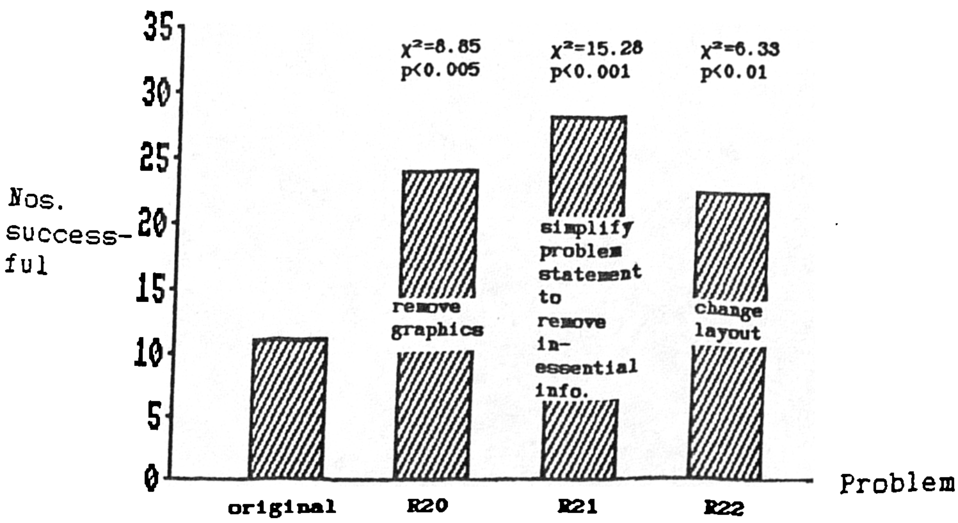
$$\begin{array}{r} 263 \\ \underline{-197} \end{array}$$

When the sum was changed to 263-197, the response was "oh, so that's 3 and 63 ... it's 66". Separating the problem to form two distinct parts produced a problem of similar difficulty to a pure subtraction sum set out according to the standard algorithm. It is worth noting that this is in contrast to the results for problem 6 where rewritten version R12 was easier than the rewritten word problems. This point will be discussed below when looking at general trends within procedural factors (Section 6.3.3) The results here, like problem 2 and 6, suggest that a number of variables affect the difficulty of a problem and that they stem from both the procedural and conceptual sphere.

Problem 12: In a garden there are
2 clumps each with 145 snowdrops,
3 beds each with 72 daffodils, and
5 beds each with 50 daffodils.
a) How many snowdrops are there?
b) How many tulips are there?
c) How many daffodils are there?
[only (a) was used]

The difficulties experienced here were mainly conceptual and these difficulties were explored in three different rewritten problems.

Figure 1f: Chi-squared tests comparing problem 12 with each rewritten version. (DF=1)



As can be seen from the above figure all the rewritten versions were statistically significant in comparison to the original version. The version where the problem statement was simplified and reduced proved the easiest. Thus a suggestion that a range of variables affect the problem difficulty, but that the need to extricate the essential information creates great difficulties. Perhaps further support for the importance of the "cognitive workbench".

6.3.2 General trend of chi-squared tests results. The trend is for an overall improvement in performance with generally higher scores on the rewritten materials than on the original problems. This could indicate that the children's understanding of mathematics has improved generally in the four to five months interval between the case study interviews and the tests using the rewritten versions. This may be the case to some extent. However, the trend towards improved performance is not uniform, and in one problem (5) none of the rewritten word problems show a statistically significant improvement. Problems 2 and 6 have two rewritten versions that do not differ from the original on the statistical tests. Thus closer investigation of the trends suggest the involvement of other factors.

(i) **Conceptual factors affecting word problem difficulty.** Three different types of variables are in evidence. They relate to the difficulty of: (a) identifying the relevant information, (b) understanding the mathematical meaning of particular words and phrases, and (c) the understanding of keyword/verbal cues.

a) Identifying relevant information. Changing the problem to make it easier to identify the relevant information improved the children's ability to reach a solution for problems 3, 9 and 12. In the case of problem 9 this change to the problem produced more correct answers than did simplifying the procedural aspects of the problem. For problem 12 this change also seemed most effective in producing an increased number of correct solutions.

b) Understanding the mathematical meaning of words. Understanding the mathematical meaning of words that are also used in ordinary English has been studied by Kane (op.cit.) and others. This difficulty was much in evidence in relation to the phrase "difference between". Many children gave what might be termed "common sense" answers to the question of "what is the difference between these numbers". The answers often stated "one is higher/lower than the other", or "one is bigger/smaller than the other". This type of problem is always treated by the textbook as a subtraction problem. However, it was noted during the task based interviews that a number of children treated this as an adding on problem and achieved a correct solution that way. For problem 6 where the original used the phrase "more than", the only rewritten version that was not significantly easier than the original was the one employing the phrase "difference between". This is an indication of the difficulty this phrase causes. The fact that a number of children interpreted this as an "adding on" rather than a subtraction sum might indicate that the difficulties the children experience with this phrase are not only a misunderstanding of the precise mathematical use of the phrase but that their common sense understanding of mathematics suggest one way of solving it: adding on, whilst the teacher insists on a different method: subtraction. Both these methods work but the former almost entirely relies on mental arithmetic and can become difficult with larger numbers. It is thus suggested that the different use within mathematical English and ordinary English of the phrase "difference between" is not the only cause of its difficulty but that there may also be a discrepancy in the child's "common sense" representation of the problem to that of the expected representation in the classroom.

c) Verbal cues or keywords. This is a much studied area, and the two terms verbal cues and keywords are used by different researchers. Here these terms are considered as the same and will be used interchangeably. The theoretical aspects of verbal cue theory is considered in Chapter 2, pp. 55-57 so that aspect will not be considered in depth this chapter. The reliance on verbal cues seems to be in evidence for some of the problems but not others.

In their 1975 study Nesher & Teubal (op.cit.) found evidence for the use of verbal cues. They suggest that when the cue indicates an operation opposite to that which is required it is particularly powerful. A later study (Nesher, op.cit.) suggested that the relationship was more complex than first thought. In this study verbal cues seemed to have had an effect in some problems, but not others. The word "more" could be suggested to indicate addition, as could the word "given". The use of "more" in problem 6 ("how many more does Billy need to make 301") could have had this effect. Changing it for "less" produced an increase in the number of correct responses that was statistically significant. However, changing the word "given" in problem 2 ("is given a start of") to a phrase less likely to suggest addition did not significantly increase the probability of a correct response. The effect of some aspects of the language of subtraction - "more than/less than/difference between" - as used in problems 5 and 6 will be looked at in greater detail below where evidence for the verbal cue theory is examined.

(ii) **Procedural factors affecting difficulty of a word problem.** The main difficulty seemed to lie with the need to carry in the mathematical operation, with carrying across zero being particularly problematic. It is interesting to note that changing the numbers to create a sum that did not require carrying produced a statistically significant improvement for problem 6 but not for problem 5, though the trend was in the same direction. For problem 6 reducing the numbers with no need to carry was no easier than removing the need to carry.

The difficulty of carrying in subtraction problems was particularly evident in the pure subtraction sums that were included. These sums were, on the whole, easier than the word problems, but not significantly so. However, when it involved an "uneven" sum, that is one number in the hundreds and one in tens only this sum on its own was no easier than the actual word problem from which it stemmed.

General trends - conclusion. It seems then that there is no single variable or a small number of variables that account for the difficulty across different word problems. Rather a variable may have a particular effect in one problem but not in another problem. In some cases it seems that the children responded to the surface structure of the problem by attending to a verbal cue but at other times they did not. Clements (1980) also emphasises this varied interaction of factors that affect the errors children make when solving word problems. He suggests they stem from two broad categories: question variables, such as syntax, level of mathematical understanding required; and person variables, such as motivation and ability. These variables are likely to interact in a varied and idiosyncratic manner.

6.3.3 Closer analysis of the language of subtraction and evidence for the use of verbal cues.

The responses to the rewritten versions of problems 5 and 6 that contained the phrases "more than", less than" or "difference between" will now be examined (see pp. 178-179). The unsuccessful responses of the rewrites have been further subdivided into "types of error". Examination of the scripts suggested the possibility of dividing the errors into two main categories: conceptual and procedural. Conceptual errors are those where the child chooses the wrong operation or is completely unable to solve the problem. Procedural errors are those where the child has seemingly identified the correct operation for the problem but is unable to carry out the operation successfully. The focus is on the problems using the phrases "less than", "difference between" and "more than". These particular phrases create a great number of difficulties and link directly to three of the variables discussed in Section 6.3.2. The use of keywords/verbal cues and the lack of understanding of the mathematical meaning of particular phrases links to the conceptual difficulties whilst procedural difficulties are indicated by errors in process skills.

(i) **Verbal cue theory.** It was shown above that changing the phrase "more than" to "less than" improved the children's performance on these problems, thus suggesting support for the verbal cue theory. However, looking at those rewritten versions of problem 5 and 6 that contain any one of these phrases the support for this theory could be questioned. In the table below the information for these problems has been regrouped into problems containing "less than", "difference between" and "more than" and is displayed in the table below:

Table 1: Responses to "less than", "difference between" and "more than" in subtraction problems

Type of Problem statement	Successful	Unsuccessful		
		Type of Error		
		Conceptual	Procedural	Total
	%	%	%	%
"Less than"	43	12	45	57
"Difference between"	43	24	32	56
"More than"	45	22	33	55

Note: "difference between" figures add up to 99% due to rounding

Verbal cue theory would suggest that "less than" should be easier than "more than". Looking at the "successful" column in the table, it can be seen that there is no significant difference of this kind. In the verbal cue theory (Nesher & Teubal, op.cit.) it is suggested that the existence of a verbal cue leads the pupil to the solution of the problem without the pupil necessarily having understood the underlying conceptual framework of the problem. The cue allows the pupil to select the correct arithmetic operation.

At this stage the results do not seem to indicate that these pupils have been relying on verbal cues to achieve a solution. Closer analysis of the types of error made by the unsuccessful respondents shows that simple verbal cues (less than) tend to produce fewer conceptual errors. This might be seen as weak evidence for verbal cue theory. However, the non-cueing phrase "difference between" seemed to be of a difficulty level similar to the cue "more than". This suggests that verbal cue theory alone does not provide a sufficient explanation of the processes at work here. It could be that the linguistic concepts underlying these types of problems are too difficult to be understood by some children of this age group. This has in fact been suggested by Lean et al (op.cit.). They suggest that many young children are "not ready for so-called 'balanced diets' of verbal arithmetic problems". The children, these researchers feel, have not developed conceptual skills suitable for the complexities that these problems involve.

6.3.4 Evidence to support the notion of a "cognitive workbench" and the effect of schemata on the solution of word problems. The discussion above on verbal cues suggests that the reliance on keywords is not consistent from one problem to another, and that other factors affecting difficulty do not have the same or a similar effect in different word problems. Britton, Glynn & Smith (op.cit.), when looking at reading processes, suggested the idea of a cognitive workbench to explain text difficulty. Essentially they suggest an area of the brain that is of limited capacity. Its function is to act as a link between external stimuli and internal knowledge stored in the form of schemata. Incoming information would be processed and interpreted here, in the case of reading drawing on elements of the text and linking this to the knowledge storage in the brain. In order to process the text items of the text will need to be stored temporarily whilst relevant internal information is found. The effectiveness of this workbench would then depend on the amount of information that needs to be accessed in an internal knowledge network. Any understanding of a particular text would depend on the complexity of the text and the availability and organisation of internal knowledge. The "workbench" is where this critical interactive encounter takes place. Britton et al suggest that this workbench is of limited capacity and as items are moved into it others will need to be cleared to provide the necessary space. The cognitive workbench idea was originally applied to reading processes. However, it seems applicable to solution of mathematics problems too, indeed to all cognitive activities where an external input has to mesh with previously stored knowledge and experience. The limited capacity would explain why the intrusion of extraneous information in a problem would make it more difficult - it might occupy valuable space on this limited capacity workbench and thus perhaps excluding or blocking essential information. This could explain the incorrect responses to problems 3 and 12 in particular. These two problems contained superfluous information, and in the rewrites the removal of this superfluous information improved the children's ability to solve the problems. If the procedures required for solution are not well established then again extra workspace would be required. This would be the case of the children who have not learnt to use the standard algorithms effectively. Each step requires a great deal of information to be sought and organised before solution can be reached. In the solution of word problems if the reading requires a great deal of effort it may interfere with the ability to select a solution. Equally if the solution requires a great deal of effort the child may lose track of what s/he is trying to find a solution to.

This workbench would be the connecting link between the text/word problem and the already stored knowledge in the brain. Many researchers suggest that knowledge in the brain is stored in the form of linked networks (a summary of this type of research is provided by Slack 1978). The creation and accessing of these networks has been widely studied and can be considered under the broad term of schema theories. The growth and elaboration of schemata in children has been investigated and described by Piaget and many others. Minsky (op.cit.) through his theory on frames suggests highly elaborate and stable schemata that affect our everyday behaviour. Frames allow us to select appropriate behaviour for a particular setting or situation without having to attend to all the particulars of that situation. Riley et al (op.cit.) suggest that, in order to solve word problems, children require certain understanding organised into schemata. Without the knowledge contained in these schemata the child will be unable to reach a solution. These schemata are of different types and need to be used in conjunction to achieve a solution of a word problem. Thus not only lack of knowledge but inability to link different types of knowledge may cause failure. The type of schemata suggested by Riley et al as being needed for the successful solution of simple word problems have been used as a basis for computer programs. These programs have been used to simulate children's problem solving methods for simple word problems. The results achieved have provided support for the ideas of Riley et al.

What must be considered, though, when looking at Riley's computer programs solving word problems and children solving the same type of problems is the difference between computers' and humans' abilities in problem solving. Computers are very effective at following clearly laid down paths and at calculating the answers but poor at coping with any deviations from the laid down rules. People are not always so good at following or remembering particular rules but can often circumvent this problem by drawing on some associated type of knowledge. So, for example, a computer program devised to solve "difference between" type problems would always employ subtraction to reach a solution; some children would follow the same path, or a similar one to the computer, whilst others would choose to add on. So the type of schemata that children build up to represent their mathematical knowledge may not be stored in exactly the same form as that of the computer and the way this knowledge is accessed could also be different. The computer programs developed are of a very limited range and have a limited knowledge base to draw. This is in contrast to children who have a vast amount knowledge stored in the brain. How this links will be depend on a large number of factors. The response to the phrase "difference between" could invoke a number of frames or schemata some of them seemingly unrelated to maths, such as: "the avoidance schema" - where the pupil employs all his/her knowledge to simply look busy, the "common sense schema/ordinary English schema" (this would result in answers such as one is bigger, larger or higher), correctly the "subtraction schema", or possibly equally correctly an "adding on" schema. Having invoked the "subtraction schema" the correct "sub schemata" or "procedures" need to be called up for a correct solution to the problem to be achieved.

Schema theory provides the means to explore both similarities and differences between children. Similarities exist between children growing up in the same culture. A number of experiences are likely to be similar for those in this culture to allow the development of a shared "knowledge base". However, individuals will also encounter some experiences that are not so universally encountered and therefore likely to produce more idiosyncratic schemata. For example, the child who frequently experiences failure at school is more likely to build up "avoidance to work" schemata than the successful child. The child who fails to build up schemata that differentiate between the mathematical usage and ordinary English usage of a number of words and phrases will also encounter difficulties. Evidence for lack of this type of schema was suggested in this research by the number of children who failed to understand the mathematical sense of "difference between" and other words and phrases. It may not be lack as such of a schema, rather that the strength of the "ordinary English" schema allows the ordinary English interpretation to be accessed more easily. Repeated success with informal algorithms, such as "add on", is likely to strengthen that type of schema and make it more difficult to replace it with a standard algorithm schema. Thus the more a schema is used the more likely it is to be remembered.

It is thus suggested that schema theory linked to the idea of a "cognitive workbench" provides a theoretical model that offers a structure and an insight into the difficulties experienced by children. How the schemata are created, elaborated and linked internally is for future research to investigate. The most effective way of helping children to create suitable and stable schemata for mathematical development should also be investigated. Clements (op.cit.) reports on an experiment where children, who had difficulty in transforming word problems into a suitable format for solution, were given special training over a period of time. Compared to a control group they performed more effectively and this improved performance was maintained over a period of time. Presumably this type of specific training allows the development of suitable schemata. The way schemata develop through experience and provide a means of interpreting the vast amount stimuli that is daily encountered by an individual is discussed more fully in Chapter 2 which looks at the theoretical background to this research.

6.4 Inter-school differences and similarities.

Table II shows the overall successful and unsuccessful responses of each school for Phase 1 and Phase 2. A ratio for both the phases will be included as this allows for comparisons between schools and between Phase 1 and Phase 2 performance.

Table II: Total number of successful and unsuccessful responses for each school for Phase 1 and 2.

School	1 S	1 US	2 S	2 US	3 S	3 US	4 S	4 US	5 S	5 US	Total S	Total US
Phase 1	50	54	41	63	47	57	31	60	16	88	185	322
Ratio	1:1.1		1:1.5		1:1.2		1:1.9		1:5.5		1:1.7	
Phase 2	121	79	91	109	98	77	111	64	106	69	527	398
Ratio	1:0.7		1:1.2		1:0.8		1:0.6		1:0.7		1:0.8	
Chi-sq. tests comparing Phase 1 with Phase 2	3.801 p<0.05 (DF=1)		0.796 Not sig. (DF=1)		2.635 Not sig. (DF=1)		19.578 p<0.000 (DF=1)		52.31 p<0.000 (DF=1)		54.151 p<0.000 (DF=1)	

Note: There were only thirteen problems including subdivided problems in Phase 1. Phase 2 contained a number of problems which were not originally considered separate but which were subdivided during the analysis. Hence the greater total number of responses in Phase 2.

Phase 1 responses are based on 8 children in each class in schools 1, 2, 3, and 5, and 7 children in school 4. Phase 2 responses are based on 8 children in each class in schools 1 and 2 only, the rest are based on responses from 7 children in each class.

6.4.1 Differences in improvement in performance between Phase1 and 2. As discussed in Section 6.3.3 there is a consistent trend for improvement in performance between phase 1 and 2. Table II shows that this improvement is particularly apparent in schools 4 and 5 where statistically highly significant improvements were found. School 1 also showed a statistically significant improvement but only at the 0.05 level. There are several possible factors that may account for this uneven pattern of improvement such as:

(a) school changes/unusual events - including change of teacher
(b) extra help for the teacher in the classroom
(c) removal of difficult pupil(s)
(d) timing of data gathering
(e) change in method of data gathering techniques

It would be best here to look at the two schools with significantly improved performances separately as there were some changes relevant to one but not the other. School 4 will be examined first.

(i) School 4.

(a) School changes: This class experienced a change in teacher after the Christmas holiday, as the previous teacher was retiring. This new teacher was experienced and seemed very keen to improve the children's experience of mathematics. She thoroughly reorganised the maths activities corner, she changed the type of work given to one very bright pupil so as not to "just give him more of the same" but to "actually make him think" (these were her own words). With the previous teacher this child had just been working further and further ahead in the SPMG textbook. She also discovered that one of the pupils that was part of the sample for this research, was lacking the necessary skills to use the formal subtraction algorithm. This child, together with one other child, received a lot of teacher attention to remedy this deficit. It was evident from the performance this child produced in phase 2 that this "remedial" work from the teacher had not been in vain. This school also experienced an HMI inspection during the period that Phase 2 was being carried out. Listening to the talk in the staffroom and seeing the work in the class that was part of this study it was clear that a great effort was being made to show up the school at its best. Thus a number of changes had occurred in this class which may have had an effect on the children's mathematical performances. Relative importance cannot be determined by this type of research which is focussed on individual children's performance on selected word problems, but would require research that looks more specifically at the different factors that affect teaching and learning in the classroom. There was no evidence that any changes had occurred relating to category (b) or (c) above. However, it is possible that category (e) - changes in data gathering techniques - may have had an effect. These changes will have affected all the classes and will therefore be discussed below in Section 6.5.

(ii) School 5:

(b) Extra teacher help: This class did not have a change of teacher, yet it showed a most significant change. This was a large class (33) with a high proportion of boys. This, according to both the headteacher and class teacher, made it rather an unruly class. The headteacher had therefore made the decision to give this class considerably more learning support help than the other classes were getting. This learning support teacher spent several mornings working with the class teacher, either by removing small groups of children or working in the classroom. This allowed for more individual attention to be given to each child.

(c) Removal of disruptive pupil: This class also had one child that was particularly disruptive. On one occasion, during Phase 1 of this study, when the class as whole was being observed this child very carefully managed to manipulate himself so as to cause his chair to fall over when he stood up. The chair was then righted and kicked over again. The teacher managed to display her annoyance without any loss of temper and was fully in control of the situation the whole time. It was clear though that this type of incident was not unusual and it naturally had a disruptive effect on anything else going on in the classroom. This child was removed from the class for a number of weeks during the spring term and it is possible that this created an atmosphere in the classroom that made work easier. It would also allow the teacher to spend more time with the rest of the children.

(d) Timing of data gathering: During Phase 1 school 4 was the last school to be visited. This meant that the children were in the process of starting division. This may have caused some confusion with the multiplication problems. Examination of the success rates on these problems (original problems 10-13) show a relatively poor performance by these children on these problems. However, the performance on subtraction problems was also poor. Subtraction is less likely to be affected by this type of confusion as early division problems always involve one number with at least two digits and one with only one digit. So, although there may have been some confusion due to the stage of learning it is unlikely to explain all the differences that were noticed between Phase 1 and Phase 2. The method of data gathering was changed. This was the case for all the children involved and will be looked at in Section 6.5.

Inter-school differences - conclusion. A number of different factors that may have affected the children's performances in Phase 1 and 2 have been examined. Their relative importance cannot be determined within the remits of this study as the focus is on children's ability to solve particular word problems. However, they constitute a significant element, and would provide an interesting starting point for future research on the role of teacher/pupil and pupil/pupil interaction in the learning of mathematics.

6.4.2 Inter-school similarities. As in Phase 1 the problems that were found to be particularly difficult were those relating to subtraction with the "more than/less than/ /difference between" causing particular difficulties across all the schools.

A number of factors were identified above that may have affected the greatly improved performances in Phase 2 in schools 4 and 5. It must be pointed out here that school 4 was not the only school to see the change of teacher, School 3 also experienced such a change. However, unlike school 4, school 3 had not only a change of teacher but a succession of three different teachers during the spring term. The situation had stabilised in the summer term with the same teacher staying on until the end of that term.

School 5 was identified as having a particularly disruptive pupil that was removed. School 3 contained several disruptive pupils, school 1 had one that was particularly disruptive. However, in none of these other classes were the pupils removed.

6.5 Changes in data gathering methods. Phase 1 involved individual interviews with children solving thirteen different problems. Phase 2 consisted of a series of paper and pencil tests over a period of three weeks. The children worked on the problems as individuals but seated in a group. No help was offered by the researcher. It may be that a number of children felt more confident working without being observed closely by the researcher. It could be that the children from two classes which showed great improvement in Phase 2 were not used to discussing and explaining their problem solutions in the way they were asked to do it in Phase 1. Phase 2 then would suit their mode of operating better. However, the children all seemed relaxed and willing to carry out the tasks in Phase I, though within each class there were some that were more forthcoming than others. It therefore seems unlikely that this type of explanation would have such a great effect as to produce the differences that were shown in Table II.

6.6 Conclusion. Phase 2 of this study aimed to investigate more closely six of the word problems used in the task-based interviews of Phase 1. This was done in order to try to identify more precisely the factors within the word problem that present the greatest difficulties. A number of rewritten versions were created and these were given to the Phase 1 sample of children as paper and pencil tests.

The responses to these rewritten problems showed a generally improved performance suggesting perhaps a general improvement in the children's understanding of mathematics. However, this improvement was not statistically significant for a number of the problems. The explanation of "general improvement", due either to general maturation and/or further teaching or to the structural changes in the problem, does not seem sufficient. Subtraction problems, particularly those using the phrases "more than", "less than" and "difference between" were still causing considerable difficulties despite the children having been taught these type of problems over the school year. Lean et al (op.cit.) were quoted as suggesting that this type of language is conceptually too difficult for many children of this age group. Procedural errors in subtraction, particularly those related to "carrying" were also noticeable. It was thus shown that there was a lack of evidence for the existence of a small number of factors that consistently influence the difficulty of a word problem.

Evidence for the use of keyword or verbal cues was examined and found generally not to explain the results. The notion of a "cognitive workbench" with limited capacity and the general organisation of knowledge into schemata was proposed to account for the responses given by the children. However, the theory of the "cognitive workbench" at this stage only seems useful as a tentative explanation of the behaviour observed. There seems no way at this stage actually to test whether in fact limited memory capacity is indeed responsible for the difficulties observed. Schema theory with knowledge organised as a set of linked networks has been suggested by many researchers. Computer programs testing the plausibility of this idea generally supports it. However, a word of caution was sounded when applying the solution processes employed by computers to explain human problem solving behaviour.

Two of the schools showed statistically highly significant improvements in the Phase 2 responses. A number of factors were examined that may account for these changes. These included change in teacher, significant events in the school, removal of disruptive pupil(s) and the timing and changes in the data gathering. However, the relative importance of these factors could only be touched upon here as this research focused on the individual child and particular problems. It would be necessary to conduct further research focusing on teacher/pupil and pupil/pupil interaction in order to illuminate the significant relationships.

This phase of the study provided a number of interesting findings. However, the effect of rewriting the problems could not be discussed with any certainty due to the time lag between the presentation of the original and the rewritten problems. The final stage of the project comprised a further study, using a different, but similar sample. This sample had not had the extensive practice with the original problems that the Phase 1/2 sample had. This study - Phase 3 - presented the original problems alongside the rewritten versions. It forms the topic for Chapter 7.

CHAPTER 7 — MAIN STUDY: PHASE 3

7.1 Introduction.

Phase 1 (Chapter 5) looked at the difficulties experienced by forty middle ability range children when attempting to solve SPMG Stage II word problems. The task-based interview was the main tool here. Phase 2 (Chapter 6) then examined the effect of structurally altering these word problems. In Phase 2 these altered problems were re-presented to the Phase 1 sample. The original problems were not presented to this sample as they had had considerable exposure to these word problems during Phase 1. It was, however, felt necessary to present the original problems alongside the rewritten versions to a similar population. This would allow the effects of the alterations to the problems to be considered in comparison to the original problems without any extraneous variables, such as time and further teaching, intruding. It was therefore decided to present the original problems and the restructured problems to a different and wider sample from a similar population. The pupils in this sample were at the same stage of their mathematics education as were the pupils in the original sample when presented with the restructured problems. This presentation of the restructured problems in conjunction with the original problems forms the basis for this chapter. Section 7.2 explains the methodology of this phase. The results then follow, together with a discussion of these results. The conclusion sums up and suggests further research.

7.2 Method.

This section will take the following form:

(i) a brief description of the overall design of this phase of the study. It will also indicate links with other parts of the project,

(ii) a description of the new sample, and

(iii) an indication of the materials used, and a comment on the organisation of the original and rewritten problems,

(iv) an explanation of the procedure involved.

7.2.1 Design. This part of the project - Phase 3 - uses the original problems and the rewritten versions. Three separate papers were created based on the nine versions that were used in Phase 2, with the addition of the original problems. These were presented to all the children from five Primary four classes in different schools over a period of three weeks. The analysis of responses looked at correct/incorrect responses. All the classes completed a standardised mathematics test. This was to allow for comparison between the previous and the present sample.

7.2.2 Subjects. The subjects for Phase 3 were the pupils of five different classes. These classes were the present Primary four classes of the schools participating the previous year in Phase 2. Thus the school data, which is contained in Chapter 5, need not be repeated. It was decided to use the whole class rather than an "average" sample.

Apart from the marking of scripts this was less time-consuming, in that the standardised tests did not have to be presented before the test papers as it was only used as a comparison between the two samples. In Phase 1/2 the standardised test had to be presented and analysed before the sample was selected, and this imposed constraints considered unnecessary for this phase of the project. For the presentation of the three test papers an effort was made to include all pupils within the class if possible, with extra visits to see pupils absent during the initial testing. However, due to time limitations, it was not possible to include those that were absent over a longer period. The numbers participating from each class were as follows (the number in bracket shows the class size):

School 1: 19 (19) School 2: 25 (28) School 3: 25 (28)
School 4: 27 (31) School 4: 30 (30)

Comparison of Phase 1/2 subjects with Phase 3 subjects. The standardised mathematics was not completed by all these children, due to absences on the day of presentation. However, it was only intended as a comparison to the previous year's sample and a sufficient number completed it to allow for such comparison. The numbers for each class completing the test were:

School 1: 19 (19) School 2: 25 (28) School 3: 26 (28)
School 4: 24 (31) School 5: 28 (30)

The results from the standardised mathematics test for Phase 1/2 (original subjects) and Phase 3 (present subjects) was divided into above average (100 and above) and below average (below 100) based on the quotients achieved in the mathematics test. Chi-squared tests were applied to this data comparing the two classes from the same school and the totals from the Phase 1/2 sample and the Phase 3 sample. No significant differences were found. It is therefore considered that these two samples are sufficiently similar for a comparison between responses to be made.

7.2.3 Materials. The rewritten versions of the original problems were used along with the original problems (see App. E). The original problems and their rewritten versions are described in Chapter 6, pp. 174-183.

Organisation of materials. In Phase 2 the rewritten problems were organised into three separate papers. In addition, each of these three papers were organised into three different versions, depending on the arithmetic operation required. In Phase 2 each child completed three papers. The organisation of the rewritten versions into three separate papers, and the subdivisions within each of these separate papers, is described in detail in Chapter 6, pp. 183-184. For Phase 3 the main three separate papers were retained - these will be referred to as the "test papers". The only change was that the original problems were also included.

The three versions of each paper were not considered necessary. Chi-squared tests on the Phase 2 responses to the different subdivisions showed no significant differences. It was therefore decided to use instead a presentation that alternated, as far as possible, the different arithmetic operations required, or expected, for problem solution. This also allowed for the most effective separation of problems that appeared similar. A problem was considered similar if only the wording of it was slightly changed (see for example, problems R5 and R6). It was dissimilar where, for example, it had been converted into a "sum" using the standard algorithm (see for example, problems R5 and R9).

7.2.4 Procedure. The three test papers and the standardised mathematics test were presented by the researcher to all the children over a period of seven weeks. School 1 started two weeks later than the rest of the schools due to teacher changes. All the children had at least one week between each presentation of the test papers. Different tests were given to children sitting beside each other to minimise copying. The standardised test used two parallel forms. The children were tested during the forenoon as it fitted in with the routine of the class.

7.3 Results and Discussion.

This section will examine the following:

(i) the overall results of Phase 3, comparing the responses to the original problems with those of the rewritten versions. The Phase 3 responses are then compared to Phase 1/2.

(ii) the general trends in the responses to the original and rewritten versions. Data from both Phase 1/2 and 3 will be used.

(iii) the total number of successful and unsuccessful responses within each individual school, giving the ratios. As the ratios are given for all three phases a comparison of performance across the phases can be made. Any significant deviations from the general trend are discussed.

7.3.1 Phase 3 responses and a comparison of responses between the Phase 2 sample and the total sample of Phase 3.

For each of the following,

- the original problem statement is given
- the responses to Phase 3 are shown in bar charts and the changes that were made to the original problem are indicated on the bar chart
- Chi-squared tests are presented to identify statistically significant differences between the original problem and the rewritten versions
- the results of the chi-squared tests are shown below the bar charts with both Phase 1/2 and Phase 3 included. Similarities and differences between the phases can then

be considered for each of the original problems and the rewritten versions

- the percentage of successful responses for the original and the rewritten versions in Phase 1/2 and Phase 3 are shown in bar charts. This allows for comment on the extent to which the original problem and the rewritten versions were considered difficult in the different phases.

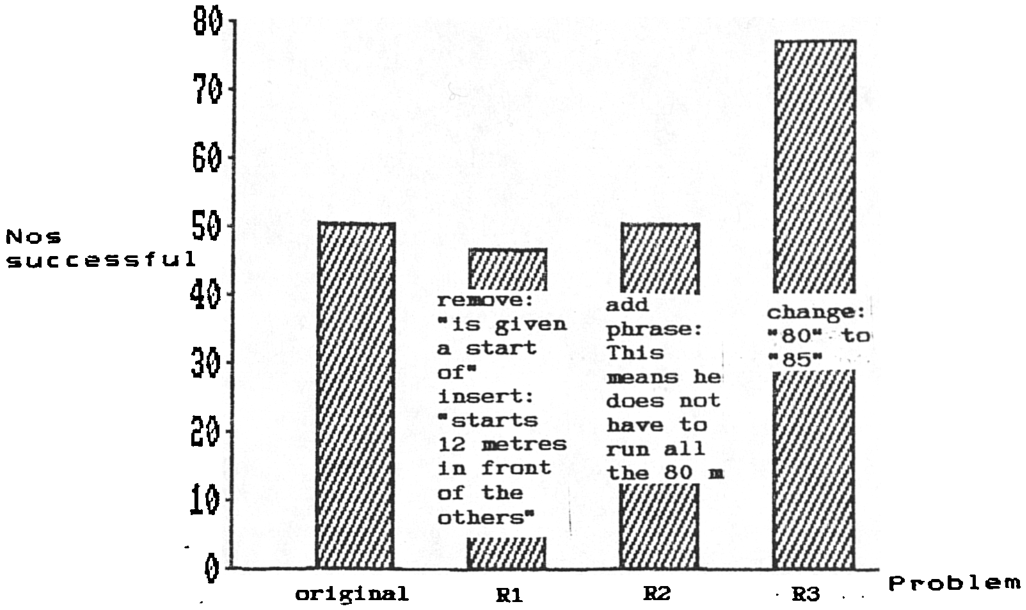
The Phase 1/2 sample was on the whole more successful during Phase 2 than during Phase 1. As there was a time lag between the presentation of the original and rewritten versions it was suggested that one possible explanation for this improved performance was general maturation and consolidation of learning. However, in the light of the ratio comparison below this does not seem to be a satisfactory explanation. The ratios show that whilst there are no significant differences between the original and the rewritten versions in Phase 3, the Phase 3 children are on the whole less successful than the Phase 1/2 sample. Thus the improvement in performance shown in the Phase 2 stage is not matched by the Phase 3 sample. Perhaps a more valid explanation is the fact that the Phase 1/2 children spent a considerable time with the researcher exploring the original problems. This may have had a positive effect on their ability to solve similar problems at a later date. These children were also more familiar with the researcher than was the Phase 3 sample as they spent considerable time with the researcher during the task-based interview in a one-to-one situation. Phase 1 was enjoyed more by the children than was Phase 2. The trend for all the problems show that the Phase 2 stage had the greatest number of successful solutions to the problems. Thus there is a possibility that the Phase 1 interviewing stage has had an effect on responses in Phase 2. This possibility will be explored in greater detail in Section 7.3.3.

Phases 1 and 2 targeted the "average" pupils, and the Phase 1/2 "average" sample had been selected by finding the average mean score for the total sample and selecting eight pupils from each class that fell within the range of +4 and -4 of this mean. A Phase 3 group of "average" pupils was identified in the same manner. The responses of these pupils were then examined to see if they differed significantly from the Phase 3 main sample. It was found, however, that the responses of this group virtually mirrored the responses of the total Phase 3 sample. Accordingly, the results for this group of pupils was not further examined.

Problem 2: Jim enters the 80 metres race and is given a start of 13 metres. How far does he have to run?

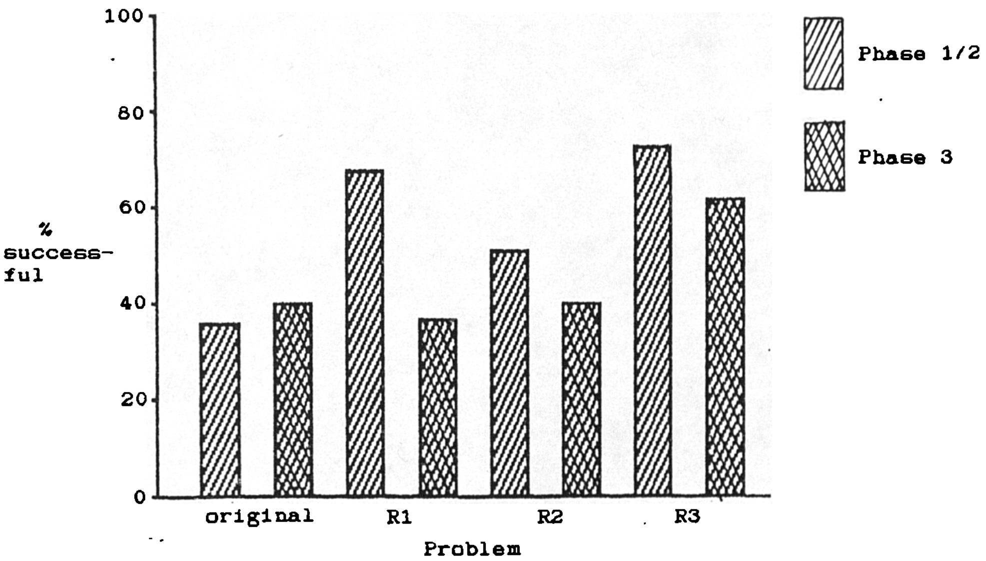
Conceptual difficulties were examined by first two rewritten versions (R1 and R2) and procedural ones in the final one (R3).

Figure 1a: Number of successful responses to the original and rewritten versions of problem 2.



Comparison of results:	Phase 2 (Nos: 37, DF=1)	Phase 3 (Nos: 126, DF=1)
original v. R1	χ^2 = 6.4 p<0.01	χ^2 = 0.2 N.S.
original v. R2	χ^2 = 1.3 N.S.	χ^2 = 0.02 N.S.
original v. R3	χ^2 = 9.1 p<0.01	χ^2 = 10.736 p<0.001

Figure 1b: Percentage of successful responses for Phase 1/2 and Phase 3.

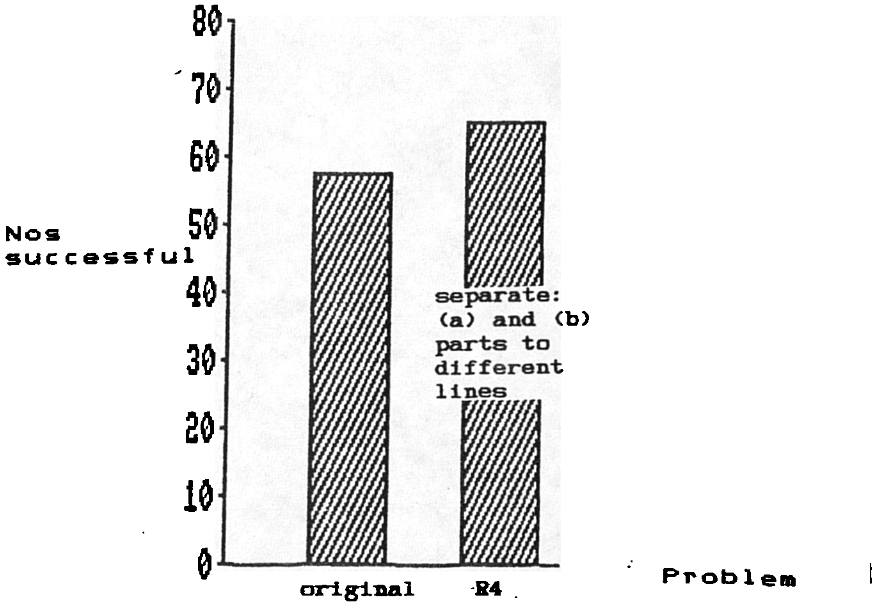


The main difference here is that rewritten version 1 was significantly easier than the original during Phase 1/2 but not during Phase 3. In fact, a smaller number of Phase 3 children were successful with R1 than they were with the original problem.

Problem 3: One morning 37 boys and 46 girls go to the library
That afternoon 39 boys and 59 girls go.
How many a) boys b) girls go to the library that day?

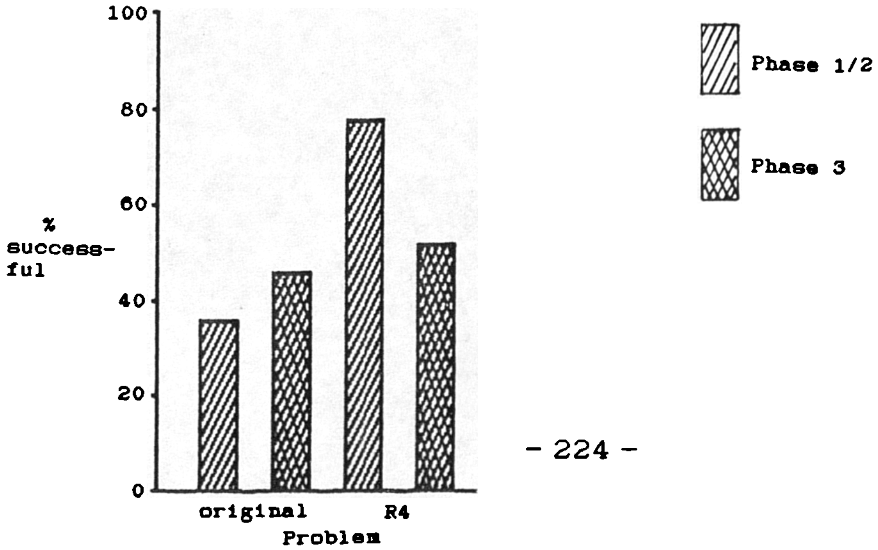
This problem had only one rewritten version which looked at the effect of changing the layout by separating the (a) and (b) subsections.

Figure 2a: Number of successful responses to the original and rewritten version of problem 3.



Comparison of results:	Phase 2 (Nos: 37, DF=1)	Phase 3 (Nos: 126, DF=1)
original v R4	χ^2 = 12.272 p<0.001	χ^2 = 0.6 N.S.

Figure 2b: Percentage of successful responses for Phase 1/2 and Phase 3.



Here there seems to be a marked difference between the responses in the earlier phases and Phase 3 as shown by the ratios above: the original was found easier in Phase 3 than in Phase 1; and the rewritten versions was easier in Phase 2 than in Phase 3. For the rewritten problem the trend is in the same direction though, with more children being successful with the rewritten version than with the original. Further analysis of the actual responses (the "workings") suggest that the presentation of the first part of the problem statement is causing confusion. Only two out of the four numbers given are required for the solution for each of the subsections of the problem. The "workings" during Phase 3 for this problem suggests that this type of presentation of data adds to the difficulty of the problem. An example of this is where a child simply adds the numbers on the top line, another is where a child proceeds to add all four numbers. A further rewritten version would have been useful here: one that reduced the information given in the "heading" statement.

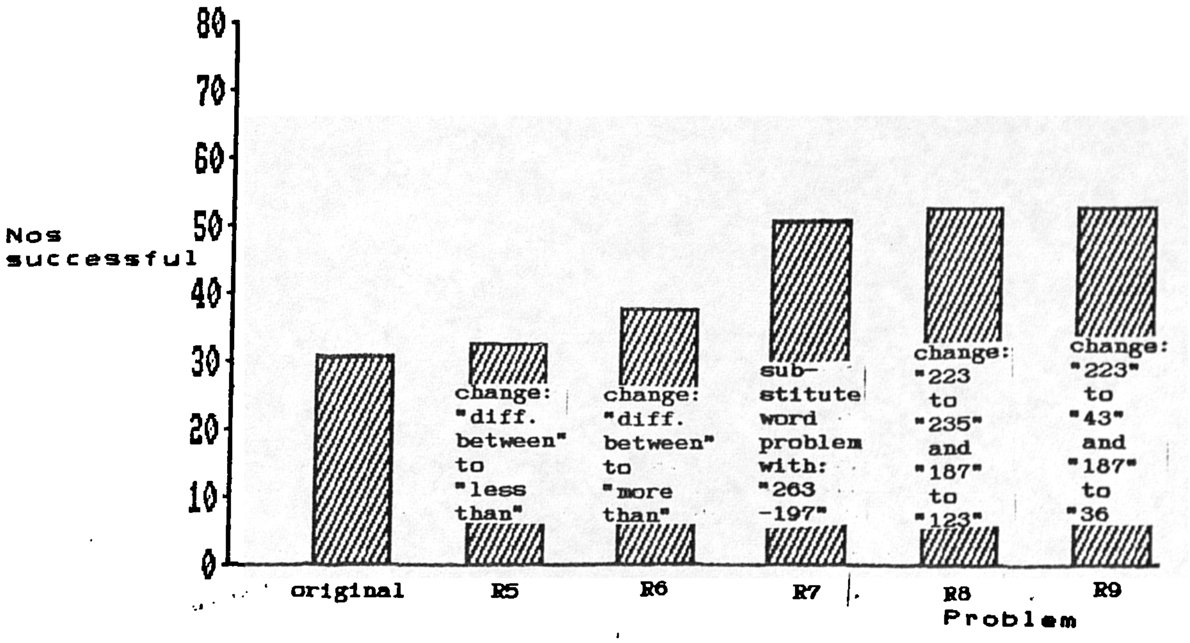
It is worth considering more generally here word problems of this particular type. SPMG textbook seems regularly to employ the technique of providing a "heading" with number information. Below this heading a selection of word problems are given that use some part of the information given in this "heading". Having to extract the correct information seems frequently to be a harder task than actually carrying out the arithmetic operation. In the case of problem 3 the required operation was the addition of two two-digit numbers. Less than half the sample managed to achieve the correct solution for the original problem, yet well over half managed to solve an addition sum that involved larger numbers and an "uneven" sum consisting of a two-digit and a three-digit number. Thus, perhaps a matter of the difficulty of information extraction overshadowing a relatively simple arithmetic operation.

Problem 5: During a game of darts Billy King had scored 187 and Jock Scott 223.

a) What is the difference between these scores?

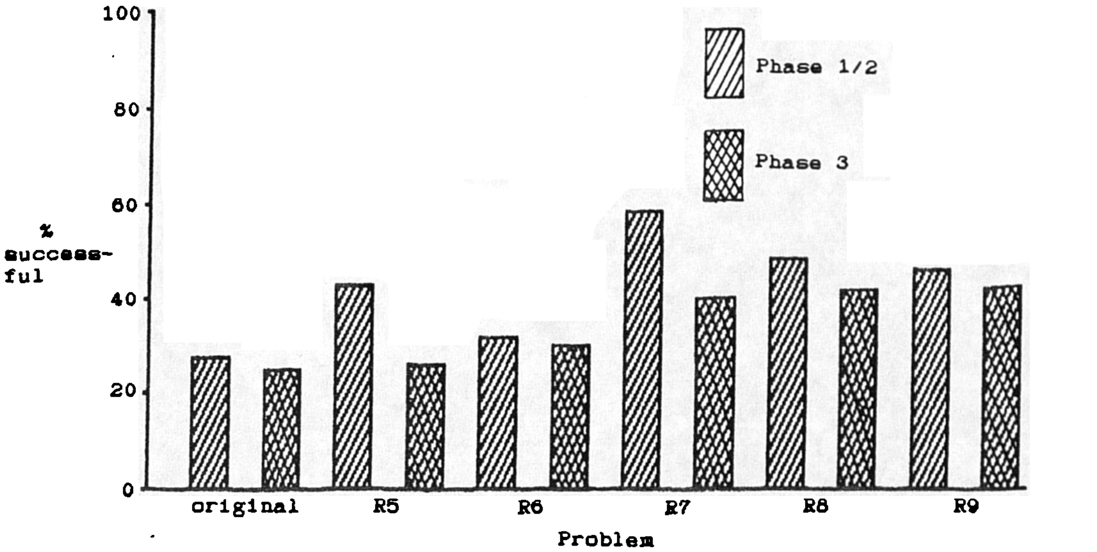
The first two rewritten versions examine conceptual difficulties (R5 and R6), the third transforming difficulties (R7), the fourth (R8) procedural ones, and the final one looks at the cognitive overload factor (R9).

Figure 3a: Number of successful responses to the original and rewritten versions of problem 5.



Comparison of results:	Phase 2 (Nos: 37) (DF=1)	Phase 3 (Nos: 126) (DF=1)
original v. R5	$\chi^2 = 1.3$ N.S.	$\chi^2 = 0.02$ N.S.
original v. R6	$\chi^2 = 0.02$ N.S.	$\chi^2 = 0.7$ N.S.
original v. R7	$\chi^2 = 6.3$ $p<0.01$	$\chi^2 = 6.5$ $p<0.01$
original v. R8	$\chi^2 = 2.6$ N.S.	$\chi^2 = 7.9$ $p<0.005$
original v. R9	$\chi^2 = 1.9$ N.S.	$\chi^2 = 7.9$ $p<0.005$

Figure 3b: Percentage of successful responses during Phase 1/2 and Phase 3



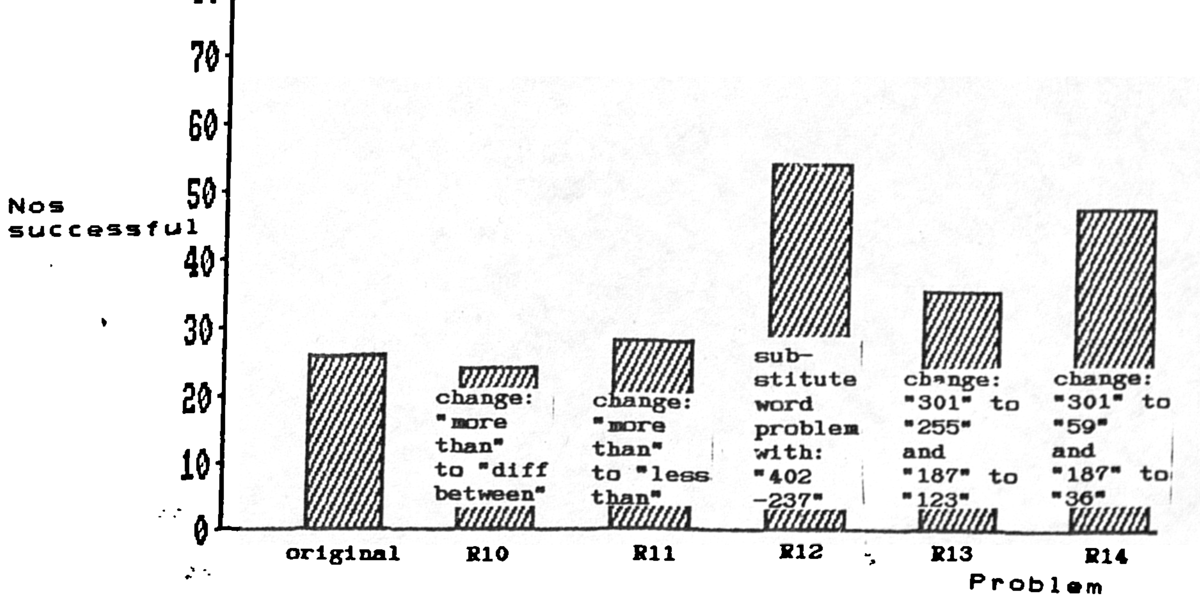
The results for Phase 2 and 3 both confirm that neither of the rewritten versions looking at the conceptual aspects of the problem were significantly easier. The changes involved removing the phrase "difference between" and substituting it with "more than" or "less than" respectively. Perhaps this is further confirmation of the difficulty of this type of language: only approximately a quarter of the children managed these types of word problems. However, this does not provide a completely acceptable explanation as Phase 3 responses to the rewritten version removing the need to carry (R8) increased the probability of a successful solution. An increase in successful solutions during Phase 3 was also noted in the "cognitive overload" rewritten problem (R9). Again the language has been retained but the numbers involved considerably reduced to below fifty. Although some of the Phase 3 responses showed the use of the standard algorithm many gave only the answer. This would suggest the use of an informal method involving a mental calculation. Although the Phase 1/2 results comparing R8 and R9 with the original are not statistically significant the value of chi-squared shows a trend signifying that these problems are easier than the original. Phase 3 which presented the original alongside the rewritten versions confirmed this trend.

Problem 6: During a game of darts Billy King had scored 187 and Jock Scott 223.

b) How many more does Billy need to make 301?

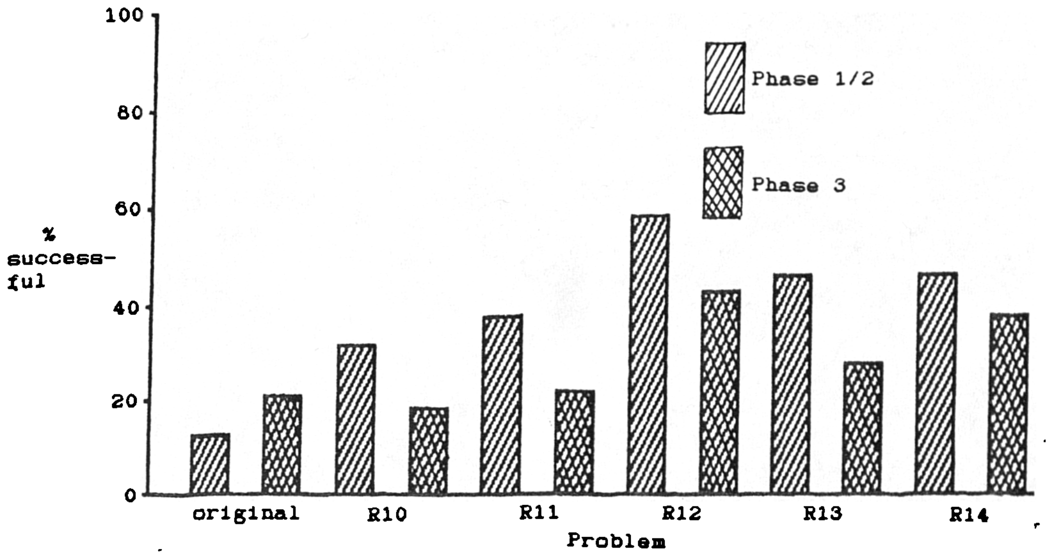
This was the (b) subsection of problem 5. Like problem 5, it had several rewritten versions. R10 and R11 dealt with conceptual difficulties and R12 with transforming. R13 examined the effect of the demands on procedural skills of carrying in subtraction, whilst R14 considered the possible impact of cognitive overload.

Figure 4a: Number of successful responses to the original and rewritten versions of problem 6



Comparison of results:	Phase 2 (Nos: 37) (DF=1)	Phase 3 (Nos: 126) (DF=1)
original v. R10	χ^2 = 3.2 N.S.	χ^2 = 0.03 N.S.
original v. R11	χ^2 = 5.1 $p<0.05$	χ^2 = 0.02 N.S.
original v. R12	χ^2 = 16.1 $p<0.0001$	χ^2 = 13.4 $p<0.0005$
original v. R13	χ^2 = 8.6 $p<0.005$	χ^2 = 1.4 N.S.
original v. R14	χ^2 = 8.6 $p<0.005$	χ^2 = 8.4 $p<0.005$

Figure 4b: Percentage of successful responses during Phase 1/2 and Phase 3.



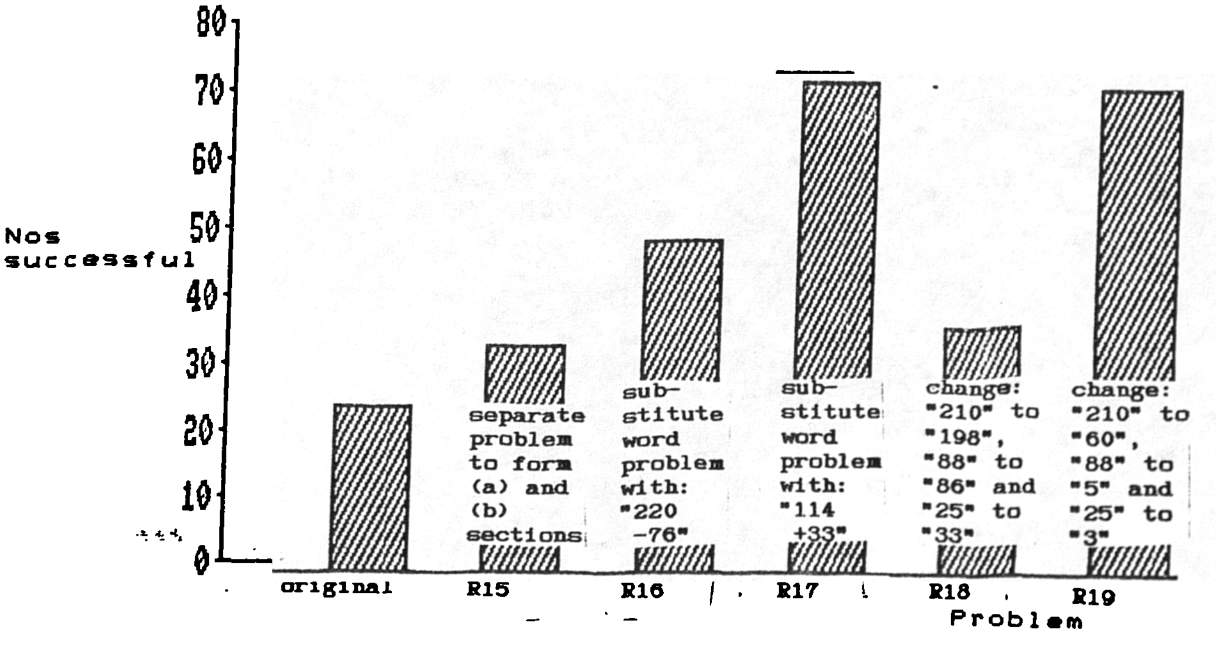
Problem 6 was the most difficult problem in Phase 1. It still created difficulties in Phase 3 but not to quite the same extent. However, the rewritten versions - R10 and R11 - that retained the original structure of the problem with alterations only in the wording: "more than" was changed to "difference between" and "less than" respectively, were more difficult in Phase 3 than in Phase 2. Less than a quarter of the sample managed to solve these problems. However, as for problem 5, it cannot be accepted that the language is not understood by all the children who failed to solve these problems. R14 which examined "cognitive overload" was significantly easier than the original, R10 and R11. Here the wording was retained but the numbers changed to be below 60. Nearly twice as many children could manage this problem.

This suggests that the relationship between variables that influence the difficulty of a problem is not straightforward. Rather it may depend on a mixture of variables having different influences in different settings. The straightforward "sum" was found considerably easier, as would be expected. Unlike problems 2 and 5 the removal of the need to carry (R13) did not differ significantly in difficulty from the original, though the trend was towards it being easier than the original.

Problem 9: Farmer Till had 210 sheep. At the market he sold 88 and bought 25. How many sheep has he now?

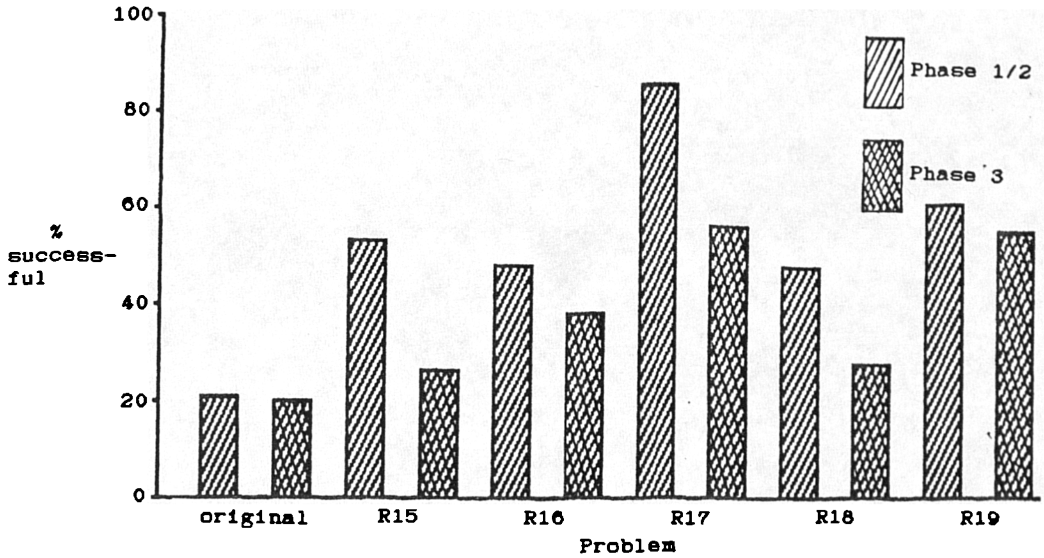
This problem had five rewritten versions. The first one (R15) dealt with layout and separated the two arithmetic operations into two subsections. Transforming was examined in two separate "sums" (R16 and R17): one for subtraction and one for addition. The final two versions (R18 and R19) examined the difficulties associated with carrying in subtraction and cognitive overload respectively.

Figure 5a: Number of successful responses to the original and rewritten versions of problem 9.



Comparison of results:	Phase 2 (Nos: 37) (DF=1)	Phase 3 (Nos: 126) (DF=1)
original v. R15	χ^2 = 7.8 $p<0.005$	χ^2 = 1.4 N.S.
original v. R16	χ^2 = 5.5 $p<0.05$	χ^2 = 10.1 $p<0.005$
original v. R17	χ^2 = 30.6 $p<0.000$	χ^2 = 35.5 $p<0.000$
original v. R18	χ^2 = 5.5 $p<0.05$	χ^2 = 0.02 N.S.
original v. R19	χ^2 = 12 $p<0.001$	χ^2 = 34.1 $p<0.000$

Figure 5b: Percentage of successful responses during Phase 1/2 and Phase 3.



As would be expected the two "straight" sums were significantly easier than the original word problem. However, the word problem where the original language was retained (R19) but the numerical operation much simplified was significantly easier than the subtraction sum (R16). Here the chi-squared value was 7.0 (DF=1) with $p<0.01$. Thus again a suggestion that the relationships between the factors that make a problem difficult are not straightforward. In Phase 2 all the rewritten versions were significantly easier than the original, in Phase 3 two out of the five versions were not significantly easier, though the trends were in that direction. Interestingly enough the version that removed the need to carry was not significantly easier. This is counter to the findings for problems 2 and 5, but in accordance with problem 6. A significant difficulty in this problem seemed to be that fact that it required two arithmetic operations to be performed. Analysis of the responses show more correct solutions for the intermediate solution (in the original this was 122) than for the final solution.

Problem 12: In a garden there are

2 clumps each with 145 snowdrops,

3 beds each with 72 daffodils, and

5 beds each with 50 daffodils.

a) How many snowdrops are there?

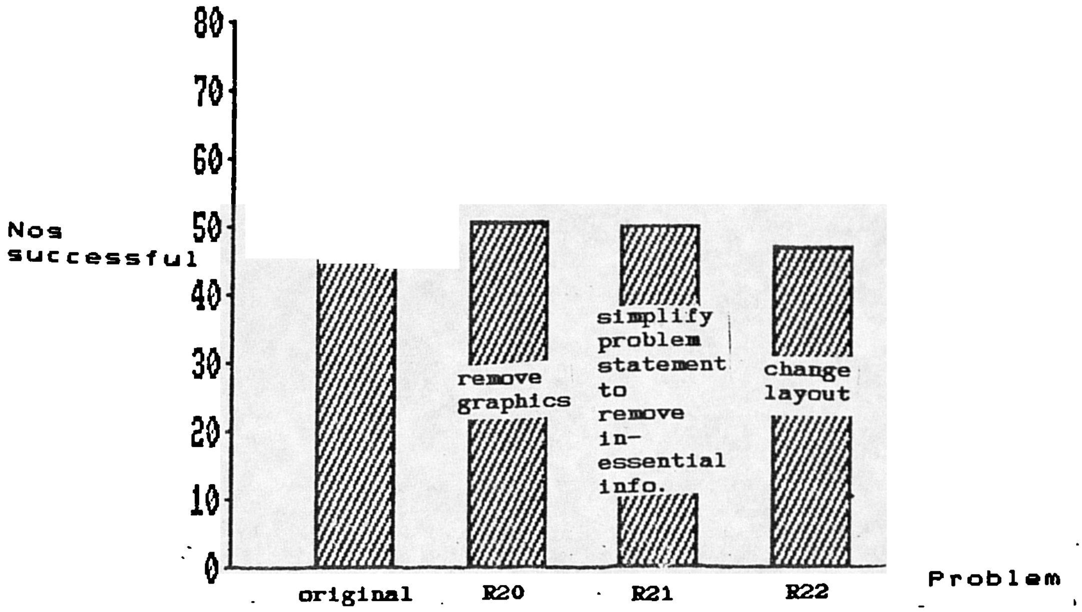
b) How many tulips are there?

c) How many daffodils are there?

[only (a) was used but (b) and (c) were retained for realistic presentation]

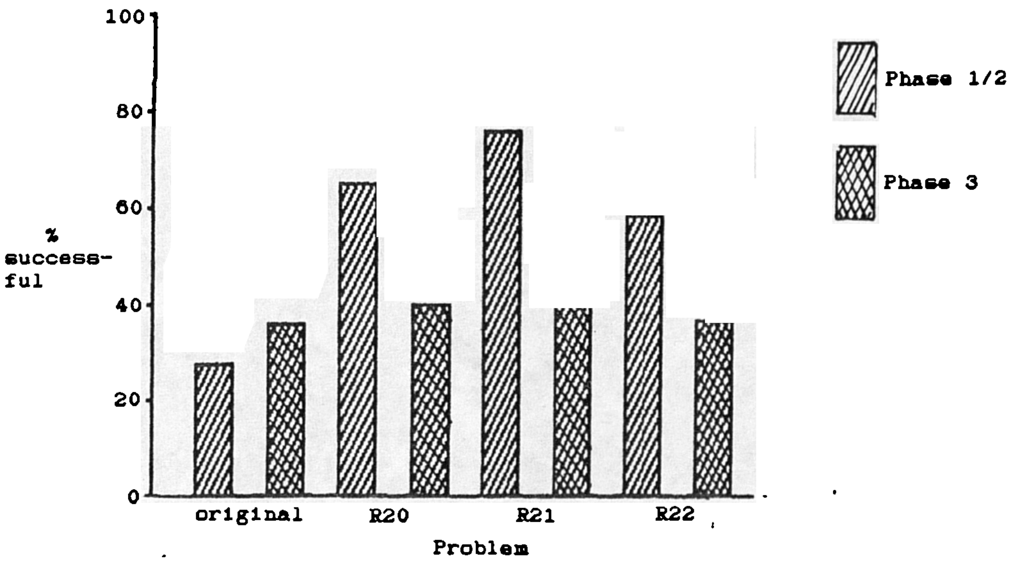
Three rewritten versions were created relating to the reading and understanding of the problem, and the use of the inessential graphics. R20 removed the inessential graphics. In R21 the problem statement was confined to the information essential for solution of (a). The final rewritten version (R22) altered the layout.

Figure 6a: Number of successful responses to the original and rewritten version of problem 12.



Comparison of results:	Phase 2 (Nos: 37) (DF=1)	Phase 3 (Nos: 126) (DF=1)
original v. R20	χ^2 = 8.9 p<0.01	χ^2 = 0.4 N.S
original v. R21	χ^2 = 15.3 p<0.001	χ^2 = 0.3 N.S.
original v. R22	χ^2 = 6.3 p<0.01	χ^2 = 0.02 N.S.

Figure 6b: Percentage of successful responses during Phase 1/2 and Phase 3.



The rewritten versions to this problem produced significant differences in Phase 2 but none in Phase 3, though the trends were consistently in the same direction. The original version of this problem was found slightly easier by the Phase 3 sample than the Phase 1 sample. However, the performance on the rewritten versions was considerably better by the Phase 2 than the Phase 3 sample.

This was numerically a very simple problem: either 2 x 145 or 145 + 145. Most children, once they had understood the language found the arithmetic simple. During the task-based interviews it was evident that the language demands of this problem were considerable. The word "clumps", in particular, seemed to have very little meaning for many of the children. However, once they had understood this word the solution came easily. It may be that this discussion of the word clumps, that many of the children of Phase 1/2 had experienced, helped during Phase 2. It had been considered, during Phase 2, when creating the rewritten version, to produce a version of this problem without using the word "clumps". However, this was not considered possible as there was no suitable substitution for this word. Thus the significant difference between the original problem and the rewritten versions during Phase 2. The Phase 3 sample had had no such extended discussion with the researcher. Lack of understanding of the "heading" part of the problem statement was shown in an emphatic manner by one child from the Phase 3 sample. He was trying to solve the rewritten version where the graphics had been removed (R20). He called the researcher's attention and complained that "there were no snowdrops". He had remembered a "similar" problem and his solution to the similar problem had been to count the snowdrops in the accompanying picture.

7.3.2 General trends across the problems.

Chapter 6 employed the division of difficulties into conceptual (comprehension) and procedural (process skills). This division will be retained here as it provides a useful way of examining the difficulties and links to the error analysis. It also allows for linkage to the discussion in chapter 6. The nature of the link between the conceptual and procedural will also be considered. As suggested by Silver (see Chapter 2, p. 42) it seems that perhaps it is the relationship between the procedural and the conceptual that is of greatest importance when investigating children's learning of mathematics.

(i) **Conceptual factors affecting word problem difficulty.** Chapter 6 (p. 193) identified three variables: (a) identifying relevant information, (b) understanding the mathematical meaning of words and phrases, and (c) understanding keywords/verbal cues. Of these the first two seem to have retained their importance in Phase 3. There seems little evidence that verbal cue is having an effect in this study. This variable has been discussed at length in Chapter 6. Evidence from Chapter 7 shows less effect of this variable and it will therefore not be considered further.

(a) Identifying the relevant information seems to remain a difficulty in problems 3, 9 and 12. Take problem 3 as an example:

In the morning 37 boys and 46 girls go to the library.
That afternoon 39 boys and 59 girls go.
How many a) boys b) girls go to the library that day?

Here the information had to be sorted into two categories: boys and girls and then the two numbers had to be added. Addition was chosen by most children, but many failed in categorising the data correctly. Problem 12 (see below) required similar action but had the added difficulty of the word "clumps". Problem 9 (Farmer Till had 210 sheep .. he sold 88 and bought 25 .. how many sheep has he now?) required a solution to be found to an intermediate sum before the final operation could be carried out. This is different from problem 3 and 12 in that there was no superfluous information that had to be ignored. However, from these three problems it is evident that identifying relevant information causes considerable difficulty in achieving a solution. It is perhaps worth noting here research that was discussed in Chapter 2 by Til Wykes (p. 26) It was suggested here that young children do not interpret informative statements in the same way as adults. A statement seemingly simple to adults requires interpretation on many levels before it has been processed so that action can be taken. Take for example problem 12:

In a garden there are
2 clumps each with 145 snowdrops,
3 beds each with 72 tulips, and
5 beds each with 50 daffodils.

a) How many snowdrops are there?
b) How many tulips are there?
c) How many daffodils are there?

Here, the complete "heading" statement has to be read before the question is reached. The child then has to work out that line 2 contains the information required for solution to (a) to be found. There is confounding pictorial information showing flowers in the margin. This information and lines three and four have to be ignored. Added to this is the difficulty of the word "clumps". Thus a complex exercise before the seemingly simple sum of 2 x 145 (or 145 + 145) can be represented in a suitable form and solved. The arithmetic operation is simple and most children during the interviewing stage were capable of doing it once they had (many with the researcher's help) penetrated the various levels of interpretation that was required for identifying the correct information.

Perhaps it could be suggested that a feature of many of the word problems in SPMG is that the level of language processing required is considerable harder than the arithmetic operation that is required. The findings of phase 3 are to some extent contrary to those of phase 2 in that a slight change in the wording has less effect. Maybe it is not the wording that matters but the levels of processing required that is of importance. In order to extract all the required information and discard some the reading needs to be slowed down. Memory demands are likely to increase as the information has to be retained for longer whilst other aspects have to be analysed. Children of this age are at different levels of development. It is therefore likely that the development of their memory capacity will also vary. For some, this task of holding different aspects of information and processing it, is a very hard task. This explanation may account for the lack of significant differences that were found between the rewritten versions of problems 2 and 12. The rewritten versions to these problems consisted only of slight changes in the wording.

(b) The difficulty of understanding the mathematical meaning of words and phrases is as evident in Phase 3 as in the previous two phases. The three phrases "difference between", "more than" and "less than" were discussed at length in Chapter 6 and this discussion will not be repeated here. Suffice to say that the responses to the Phase 3 problems confirm this difficulty. Out of a sample of 126 only twenty-six could successfully solve problem 6, for problem 5 the figure was thirty-one [these two problems asked for "the difference between" two scores (5), and "how many more" were required to reach a certain score (6)] However, here perhaps evidence of the complex relationship between the conceptual and procedural aspects of problem-solving. When the procedural demands of these two problems were reduced there was a significant increase in the number of children able to understand this language from the mathematical point of view. It could be that the numbers involved in this sum allowed the children to apply their own informal, possibly "adding on" method. Perhaps the original problem was recognised by the children as "one done in class" and therefore it made it one that had to be done "the way the teacher says", i.e. using the standard subtraction algorithm. This may have the effect of increasing problem difficulty for those children insecure when using the standard subtraction algorithm. The validity of these explanations cannot be ascertained in this project. However, the effect of teacher imposed standard algorithms and the possibility of making greater use of children's informal understanding within the classroom could usefully be explored in future research.

(ii) Procedural factors. Again the main difficulty seems to be the inability to "carry" correctly, when using the standard subtraction algorithm. The "uneven" sum was slightly more difficult than the other two subtraction sums, but not significantly so. The pure subtraction sums were significantly easier than the original problems to which they were related. This could perhaps be taken to suggest a greater procedural than conceptual competence in the children who failed to solve the original problem but managed to solve the subtraction sum. However, for problem 9 the rewritten version examining "cognitive overload" was significantly easier ($\chi^2=7$, DF=1, $p<0.01$) than the pure sum. This suggests that a greater proportion of children can understand the conceptual aspects of this problem in certain circumstances. For problem 5 and 6 no such statistically significant difference was found between the pure sum and the cognitive overload version. Here they were very much of a similar difficulty (the number successful in problem 5 were fifty-one and fifty-three respectively; for problem 6 they were fifty-four and forty-eight).

General trends - conclusion. There were some differences between the responses of Phase 2 and Phase 3: some rewritten versions that were significantly easier than the original in Phase 2 were not so in Phase 3. However, the trend towards the rewritten versions being easier was also found for all but two rewritten versions. It has been suggested that some of the improvements shown in Phase 2 were not due to changes in the structure of the rewritten problems but possibly due to in depth exploration of the problem that the Phase 1/2 sample experienced during Phase 1. However, there were some significant differences between the original problems and some of the rewritten versions during Phase 3. These are mainly that the problems examining cognitive overload are significantly easier than the original, and that they are as easy as or easier than some of the pure sums. This suggests that more children than would be expected from evidence on the original problems have some conceptual understanding of the language and the relationships involved in these problems. It is suggested that this highlights the complex interrelationship between conceptual and procedural factors. It also re-emphasises the statement made in Chapter 6 (p. 24) that the effects of the different variables interact in a varied and idiosyncratic manner.

7.3.4 School data.

The individual school data for Phase 3 is displayed in Table I showing the number of successful and unsuccessful responses to the original problems and their rewritten versions. This allows for discussion of the effect of different problem type within individual classes.

Table II shows a comparison across the phases of unsuccessful compared to successful responses in the form of ratios. Similarities and differences within individual classes across the phases are then discussed.

TABLE I

Number of Phase 3 children successfully completing the original and rewritten versions compared to those unsuccessful.

School (nos/class)	1 (19)		2 (25)		3 (25)		4 (27)		5 (30)		Total (126)	
Problem No	S	US	S	US	S	US	S	US	S	US	S	US
2 (sub)	13	6	10	15 †	1	24	15	12	12	18 †	51	75
R1	10	9	10	15	4	21	8	19	15	15	47	79
R2	11	8	9	16	3	22	13	14	15	15	51	75
R3*	13	6	18	7 †	5	20	20	7	22	8 †	78	48
3 (add)	10	9	12	13	3	22	14	13	19	11	58	68
R4	11	8	14	11	7	18	15	12	18	12	65	61
5 (sub)	7	12	6	19	1	24	8	19	9	21	31	95
R5	8	11	7	18	0	25	8	19	10	20	33	93
R6	11	8	6	19	1	24	10	17	10	20	38	88
R7*	11	8	11	14	2	23	13	14	14	16	51	75
R8*	13	6	8	17	2	23	11	16	19	11	53	73
R9*	11	8	14	11	1	24	14	13	13	17	53	73
6 (sub)	5	14	3	22	0	25	7	20	11	19	26	100
R10	6	13	2	23	0	25	6	21	10	20	24	102
R11	7	12	2	23	2	23	8	19	9	21	28	98
R12*	9	10	6	19 †	4	21	17	10	18	12	54	72
R13*	7	12	6	19	1	24	7	20	14	16	35	91
R14*	8	11	14	11 †	3	22	10	17	13	17	48	78
9 (sub/	6	13	4	21	0	25	8	19	7	23	25	101
R15 add)	6	13	6	19	2	23	12	15	8	22	34	92
R16*	9	10	9	16 †	3	22	11	16	17	13	49	77
R17*	12	7	14	11	11	14	15	12	20	10	72	54
R18	6	13	10	15	1	24	9	18	10	20	36	90
R19*	10	9	18	7 †	5	20	16	11	22	8	71	55
12 (multi)	6	13	14	11	3	22	10	17	12	18	45	81
R20	10	9	12	13	3	22	11	16	15	15	51	75
R21	11	8	11	14	5	20	8	19	15	15	50	76
R22	9	10	12	13	3	22	11	16	12	18	47	79
Total	256	276	268	432	76	624	315	441	389	451	1304	2224

* indicates rewritten version that was significantly easier than the original as measured by a chi-squared test.

† indicates significant ($p<0.05$) interproblem differences on individual school score (see discussion below)

Individual school data in Phase 3 sample. These show that on the whole all the classes follow a similar pattern with problems found difficult by one school also being found difficult in the rest of the schools. There were some differences, however, with School 3 and School 1. School 3 follows the pattern but has, overall, far greater difficulties than the other schools. School 1 deviates from the pattern in the responses to the original problem 2 and the rewritten version R3. The other four classes all find R3 significantly easier, School 1 does not. School 2 shows an interesting pattern in problems 6 and 9 in that it is the only school where there is a statistically significant difference between two rewritten versions: R12 and R14; and R16 and R19:

R12 v R19 - $\chi^2 = 4.08$, DF=1, $p<0.05$

R16 v R19 - $\chi^2 = 5.15$, DF=1, $p<0.05$

In both cases the "cognitive overload" versions are easier than the straight subtraction sum. This ties in with the findings from Phase 1 that several of the children in this school that were successful with these types of problem used their own informal method, and added on instead of subtracted. The primary 4 of Phase 3 had the same teacher as those of Phase 1/2. Perhaps this is an indication that teacher variables are affecting these children's development: the use of idiosyncratic methods are at the expense of learning to use the standard subtraction algorithm effectively. Further support for this could perhaps be taken from the fact children from this school also found R3 significantly easier than the original problem 2 ($\chi^2 = 3.98$, DF=1, $p<0.05$). In R3 the need to carry in subtraction had been removed. School 5 also showed a statistically significant difference between problem 2 and R3 (χ^2 = 5.5, DF=1, p<0.05). However, in line with the rest of the schools this school showed no other such significant differences.

TABLE II

Total number of successful and unsuccessful responses for each school for Phase 1, 2 and 3.

School	1		2		3		4		5		Total	
	S	US	S	US	S	US	S	US	S	US	S	US
Phase 1	50	54	41	63	47	57	31	60	16	88	185	322
Ratio	1 : 1.1		1 : 1.5		1 : 1.2		1 : 1.9		1 : 5.5		1 : 1.7	
Phase 2	121	79	91	109	98	77	111	64	106	69	527	398
Ratio	1 : 0.7		1 : 1.2		1 : 0.8		1 : 0.6		1 : 0.7		1 : 0.8	
Phase 3	256	276	268	432	76	624	315	441	389	451	1304	2224
Ratio	1 : 1.1		1 : 1.6		1 : 8.2		1 : 1.4		1 : 1.2		1 : 1.7	

General trends across the phases. Looking at the total number of successful responses the ratios show that there is close agreement between Phase 1 and Phase 3. A chi-squared test between phase 1 and 3 shows no significant differences. Examination of the data across all three phases, however, reveals three prominent features that demand consideration: (i) the performance of school 5 in Phase 1; (ii) the general improvement in performance during phase 2; and (iii) the performance of school 3 in Phase 3.

(i) The difference in performance in School 5 during Phase 1 and the significant improvement made here during Phase 2 was discussed at length in Chapter 6 (pp. 209-210).

This discussion will not be repeated here. It is interesting to note though, that the Phase 3 class from this school had the same teacher as the parallel class in Phase 1/2. In the opinion of this teacher the Phase 3 class had fewer difficult children. The child that had caused the greatest difficulties in this class during Phase 1/2 was now in a special school.

(ii) The general improvement during Phase 2 was commented on in Chapter 6 (p. 211) and in Section 7.3.2 (p. 7) of this chapter. The ratios for Phase 2 totals indicate that the Phase 1/2 sample was generally more successful during Phase 2. This difference is emphasised by a chi-squared test comparing the three phases - $\chi^2 = 9.3$, DF=2, $p < 0.01$. It had been suggested that maybe the improved performance was due to maturation and consolidation of learning. However, the results of Phase 3 suggest that this is not necessarily a satisfactory explanation: the responses during Phase 3 do not indicate that the children perform more effectively on class taught material towards the end of the session. In fact, the indications are that the children are no more effective in May/June than they were in October/ November. It seems that a more satisfactory explanation of the improvement during Phase 2 is the amount of time that the children spent, with the researcher, exploring the original problems during Phase 1. During Phase 1, the children were not only asked to solve the problems, but they were helped to find a solution if they were unable to do so on their own.

Comparison across the phases assumes that the two samples are drawn from a sufficiently similar population. The standardised mathematics test indicated that this was the case. The actual test scores on this test were slightly higher for the Phase 3 than for the Phase 1/2 sample. This, it has been suggested was due to time of year of testing. However, applying quotients suggested a similarity between the two populations. It is perhaps difficult to reconcile the fact that the standardised test showed an improvement in performance whilst the classtaught material showed no such overall improvement. Many of the tasks involved in the standardised test were considerably simpler than those of the project tests. Thus perhaps an indication that the majority of the children had some grasp of the most basic aspects of arithmetic but that some aspects of the word problem type material of SPMG had not been fully understood by many.

(iii) Performance of school 3 during Phase 3. This class showed a considerably poorer performance on the researcher-created tests than did the Phase 1/2 class from this school and the other classes. There are a number of factors that may have been influential - three considered of possible importance will be discussed here: teacher changes, class size and children with learning difficulties.

Teacher changes. The Phase 1/2 class experienced a change of teacher at the beginning of the Spring term. The Phase 3 children had no such change. The teacher of the Phase 3 class was an experienced infants teacher, but on her own admission had to make considerable changes to her teaching style when she moved to teaching a Primary 4 class. She also commented on her lack of familiarity with the SPMG Stage 2 material. However, in observing her dealings with the class she seemed highly competent and in control. The latter is mentioned as this is a very difficult school to teach in with a much higher than average proportion of children from difficult backgrounds.

Class size. The Phase 3 class consisted of twenty-eight children, four more than the Phase 1/2 class.

Children with learning difficulties. Both the classes involved from this school had several children with learning difficulties. The standardised mathematics test showed that both the Phase 1/2 and 3 classes had three children with a quotient below 80. However, the Phase 3 class had a larger number (nineteen compared to twelve) of children with quotients below average. There were several children with poor reading ability in this class - this was specifically pointed out to the researcher by the learning support teacher. The implications for the class teacher of large numbers of poor readers can be considerable. Many tasks set in the classroom including mathematical ones require a certain standard of reading. Without this standard of reading the teacher has to spend considerably more time explaining to the children what the task involves and there is a limit as to the type of task that can be used by the teacher. Maybe the combination of a larger class with more below average pupils has made this a more difficult class to teach than the Phase 1/2 class. It has to be considered also that the children who are not considered to have learning difficulties as such may nonetheless suffer in this type of class. They are far more likely to have less time spent on them by the teacher. The generally poor performance of this class on the three tests suggests that this may be the case. The researcher-created test material related directly to classroom taught material.

Those were the most obvious features indicated by the responses shown in Table II. Undoubtedly there is variation between the schools. These are likely to be due to a number of factors, some of which have been discussed in the preceding section and in Chapter 6. It was not the intention of this project to study teacher and school variables and therefore this topic will not be pursued further here.

7.4 Conclusion.

Phase 3 aimed to examine the effect of presenting the original word problems alongside the rewritten versions. This was done in order to examine more closely the effects of the structural changes made to the original problems. The original problems were therefore presented in conjunction with the altered problems to a new sample of Primary 4 children.

The sample for this phase of the project constituted the present Primary 4 in the schools that participated in the previous year. The same standardised mathematics test was given to this group of children as was given to the group from Phase 1/2. Quotients were calculated for the two samples and and a chi-squared test determined that there were no significant differences between the two samples. It was therefore considered valid to make tentative comparisons between the Phase 2 and Phase 3 results.

The responses to the problems, original and rewritten, showed some significant differences. Some of these confirmed the results of Phase 2, whilst some did not. Slight changes in wording or layout as were done in Problems 2, 3 and 12 seem to have had little effect on the success of the Phase 3 sample. It was suggested that the improvement shown by the initial sample in Phase 2 were more likely to be due to researcher induced factors than general maturation. Thus, it seems that structural changes of this kind are not likely to make problem solving easier. However, there were two areas where difficulties were experienced by many children in all the phases: (i) comprehension of the phrases "more than", "less than" and "difference between", and (ii) in the ability to carry in subtraction sums.

(i) Comprehension of "mathematical" phrases and the effect of changing arithmetical demands.

The original word problems 5 and 6 suggested great difficulties with these problems. However, by introducing a "cognitive overload" version of this problem it was shown, by the increase in correct responses, that a greater number of children could understand this language. It was shown that when the numbers involved were greatly reduced solution was easier. It was proposed that this was partly due to the fact that the children could rely on mental methods for the solution of the problem. This mental method it was suggested would involved adding on rather than subtraction. Thus there may be a clash between the way many children would prefer to represent this problem and the way that is expected by the teacher. So, simply suggesting that children don't understand the language of subtraction is not a satisfactory explanation for failure in solving these type of problems. Rather what needs to be explored is just when they show evidence of understanding and when this understanding seems to be lacking. It is possible that this is linked to the way that children are expected to represent the problem. In a sense a clash between an informal method using adding on and a formal method relying on subtraction is perhaps in evidence.

(ii) The suggestion above that many children find the standard subtraction algorithm difficult is emphasised by the difficulty created by problems that require "carrying". Many of the children showed limited understanding of the formal subtraction algorithm. When the need to carry was removed the problem became significantly easier (except in the case of problem 9).

Thus there were two areas where structural changes were made and shown to have a significant effect. They are seemingly separate in that one seems to stem from the conceptual domain and the other from the procedural. However, a link can be made in that difficulties in the procedural domain seems to have an affect on the interpretation of language. When the language can be interpreted and represented using informal methods success is more likely for some children. Again perhaps an indication that Silver's (op. cit.) suggestion is correct. He felt that the domains should not be examined in isolation but rather that it is the interrelationship between the conceptual and procedural that needs clarifying.

The relationship between children's informal methods and their understanding of subtraction would be well worth exploring further. If a way could be found to help more effectively the many children who seem to find subtraction an insoluble riddle it would be a valuable advance in mathematics education.

A number of variables outwith the problem structure were explored in an attempt to account for variability in performance between school 3 and the rest of the sample in Phase 3. It was suggested that school variables worth considering here were teacher changes, class size and the number of children with learning difficulties in the class. These variables have not formed part of this project and could therefore not be pursued in depth. Chapter 6 also touched upon, in a very limited manner, factors outwith the school that may be influential. It is clear that all these factors have an influence. However, this type of study is too limited to offer anything more than tentative suggestions as to which variables may be affecting the variability in performance that was shown in Phase 3.

CHAPTER 8 – OVERVIEW, REFLECTIONS AND CONCLUSION.

8.1 Introduction.

This project set out to explore how children cope with the word problems they meet with at school. The motives for the project stemmed from the researcher's awareness, from the literature and the classroom, of the many difficulties that a number of children seem to experience with word problems. The classroom provided the setting for the project, which consisted of four phases:

- the exploratory study and Phase 1 of the main study, using mainly qualitative research methods. The main aim of these parts of the project was to explore and develop the use of task-based interviews as a means for examining children's difficulties when solving word problems, and for considering these difficulties.

- Phase 2 and Phase 3 of the main study, using mainly quantitative methods. The main aim of these parts of the project was to investigate whether structural changes to the word problems affected the ability of the children to solve the problem. These two phases built on the earlier ones by using the difficulties evidenced in Phase 1 as a guide for rewriting the problems.

This chapter discusses these four phases and the overall findings. Chapter 3 discussed research methods in educational research and explained the methods chosen for this project. The choice and use of particular methods affects the final results. A discussion of the chosen methods, alternative ways of exploring children's word problem understanding and the generalisability of the findings therefore require consideration. The concepts used to explore the findings need to be examined as they have an effect on the analysis offered. The implications of this research on current theories and educational practice is then discussed. The conclusion sums up and gives an overview of the sections of this chapter. The format adopted is thus:

(i) a discussion of the aims and results of the four phases of the project and the overall findings.

(ii) an examination of the methods used, including alternative ways in which children's understanding of word problems could have been explored. The generalisability of the findings are considered here.

(iii) a discussion of the usefulness of the concepts employed to explain the children's understanding. Linked to this is the problematic nature of accessing children's learning and the use of actual performance to investigate learning.

(iv) the implications of this project for theory, and its possible effects on educational practice are discussed. The role of word problems, which was considered in Chapter 1, is reconsidered here.

(v) conclusion.

8.2 Summaries of the four phases.

8.2.1 The exploratory study.

This phase of the project aimed to test out the feasibility of using a task-based interview in the classroom setting. This aim was achieved. The task-based interview coupled with an error analysis has provided useful insights into primary 4 children's problem solving behaviour.

The task-based interviews. The rationale behind this research method has been discussed in chapter 3. For this phase of the study the tasks consisted of sixteen word problems from SPMG Stage 2 (see App. A). These word problems were all from pages that had already formed part of the class teaching during the year. The interviews were recorded on tape and transcribed immediately after the interviews.

Sample. At this stage only one class was involved. Fifteen pupils from different ability groups took part in the task-based interviews. The exploratory study was carried out during the month of March. A subsidiary aim of the exploratory study was to consider the narrowing of the focus of the study to a more specific ability range within the ordinary class.

Error analysis. In order to gain further insights into the data gathered in the interviews an error analysis was adapted from that of Watson (op.cit.). This error analysis made it possible to classify errors on a conceptual/ procedural basis. It was found that not only those identified as slow learners by their teachers experienced difficulties with these word problems. Many of the children in the middle ability range also found this previously taught material difficult. It was therefore decided to narrow the focus of main study to pupils who could be considered of "average" ability.

8.2.2 Phase 1 of the main study. This phase of the study built upon the exploratory study and was carried out the following session during the period October to December. It provided further data on a larger number of pupils. The main aim was to investigate further the difficulties experienced by "average" children when attempting to solve word problems. The sample was larger and came from several different schools. It strengthened and illuminated the findings of the exploratory study. The task-based interview was used in conjunction with the revised error analysis. The number of tasks - word problems - were limited to thirteen. The word problems involving division were dropped. This was due to the fact that Phase 1 was carried out earlier in the school year than the exploratory study, and some of the classes involved had not reached the division learning stage.

It differed from the exploratory study in two important ways: the study was extended to include five Primary 4 classes; and the focus was narrowed to eight "average" ability pupils within each of these classes. This provided a total sample of forty children for this phase of the project.

Extension of sample group. By extending the study to include five schools it was possible to examine whether the difficulties that were experienced by many of the children in the exploratory study were also common to similar children in other classes. It was thus possible to build up a picture of the type of problems that were considered particularly difficult by many children in several schools. It also allowed for comment on possible other external variables which might affect problem solving success.

Narrowing of focus. A standardised, commercially produced mathematics test was used to select the sample. A sample was selected that fell within -4 and +4 of the mean for the total population.

Task-based interviews. These followed the format of the exploratory study. The children enjoyed doing the tasks and tried hard to complete them. The suggestion made, based on the experience of the exploratory study, that the children should not be helped to a solution, was abandoned after the first two interviews in Phase 1. It was considered important that the children be helped over any impasse in the early part of understanding the problem. If this was not done it was impossible to assess whether the children possessed the procedural skills required for the problem.

One very interesting aspect of the task-based interview was that it allowed for the observation of the strategies used by the children for problem solution. Throughout this project the terms "informal"/"formal" and "strategies" have been used. The implications of choosing to use these particular terms will considered in Section 8.4. However, for the remainder of this section these will be used. The word problems used had already been part of the children's classroom lessons. The textbook had very clear expectations as to how the problems should be solved. This is shown by the headings at the top of the page, and the other exercises on the page. The teachers' handbook contained plenty of advice on how to teach the relevant standard algorithm. Despite this, alternative strategies were in evidence. This was particularly so for subtraction problems. The most commonly found alternative strategy was that of "adding on". Here the child would count up from the lower number until the higher number was reached. The calculations were all done mentally. It seems on reflection and further consideration that the confusion expressed by some children when trying to select a suitable operation for these types of problems may not have been due to lack of understanding of the underlying structure. Rather the difficulty may have been related to a conflict between what seems to be the more informal, intuitive method of adding on and the school-taught subtraction algorithm.

Error analysis. Further refinements were made to the error analysis. There was differentiation between those children requiring a little help towards the solution and those requiring a great deal. The error analysis provided a means to consider the types of difficulties experienced and to link this to the theoretical debate on conceptual and procedural knowledge. It also provided a basis for rewriting some of the problems found most difficult by the children involved in the study.

It was suggested by Watson (op.cit.) and by this researcher that the error analysis may be useful for a teacher who wishes to explore individual children's difficulties. This, it is felt, is still the case. However, it does seem unlikely that it will be used in this manner by teachers. The present conditions in Scottish classroom seem to allow very little time for this type of exploration of individual children's difficulties.

8.2.3 Phase 2 of the main study. This study built on Phase 1. Its aim was to investigate the success of the Phase 1 sample in solving a number of rewritten and structurally altered word problems. A small number of word problems were selected, altered and re-presented to this sample. This phase of the project was carried out in May.

Rewritten problems. Structural alterations were made to six of the original SPMG word problems. These alterations were based on the most common type of errors, as shown by the Phase 1 error analysis, for the problems. Three main paper and pencil tests were created. These were presented to the children, one per week, over a period of three weeks. Thirty-seven of the original sample of forty completed these tests.

Results. A general improvement in the ability to solve most of the rewritten problems compared to the original ones was found. Several of these improvements were statistically significant differences. However, due to the fact that the original problems were not presented at the same time as the rewritten versions, the effects of the structural alterations could not be properly assessed. The original problems had not been re-presented to this sample as they had had considerable practice doing these problems.

8.2.4 Phase 3 of the main study. This study provided an opportunity to present the original problems alongside the rewritten versions to a different, but similar sample. The reason for not re-presenting the original problems alongside the rewritten versions is explained above. However, it was felt that in order to assess more reliably the effect of the structural alterations to the word problems it was essential to present the original and rewritten problems within the same tests. Thus the same tests were used here as in Phase 2, with the addition of the original problems. It took place in May, a year later than Phase 2.

Sample. This consisted of the five primary 4 classes. The same schools that were involved during Phase 1/2 agreed to participate. The classes were the present primary 4 classes of these schools. The same standardised mathematics test was administered to these pupils. Quotients were used to compare the sample of Phase 1/2 and Phase 3. No statistical significant differences were found. Thus, it was felt that, with caution, the results from Phase 1/2 and Phase 3 could be compared.

Results. It was found that the Phase 3 sample did not perform as well as did the Phase 1/2 sample in Phase 2. It was therefore suggested that the general improvement shown between Phase 1 and 2 was not likely to be due to maturity or greater understanding due to the teaching in the intervening period. The improved performance, shown during Phase 2, was perhaps more likely to stem from the amount of exposure to the original problems that the Phase 1/2 children experienced during the task-based interviews. However, some of the rewritten problems were still found, by the Phase 3 sample, to be significantly easier than the original ones. Of particular interest seem to be the problems involving subtraction. Initial analysis might suggest that the language was too difficult for many children. However, the rewritten versions examining "cognitive overload" suggests that this explanation is not sufficient for all cases. This aspect of the results will be discussed further in Section 8.2.5, which examines the overall findings of the project.

8.2.5 Overall findings of the project. This research project found that many children, who could be considered to be within the average ability range, experienced difficulties with recently taught material. These difficulties seemed particularly apparent in word problems where the expected solution procedure involved subtraction. Along with other terms used in this project, the use of the word "difficulties" and the term "cognitive overload" are examined in Section 8.4. These terms are retained for the discussion in this section.

Complexity in relationship between variables that affect problem difficulty. It has been suggested, both in chapter 6 and 7, that this project has not found a small number of variables that consistently affect problem difficulty across problems. Rather the relationship between variables is complex and is affected by the interpretation put upon the problem statement by the reader. To exemplify, problems 5, 6 and 9 will be used. These three problems all involved subtraction. Problem 9 also involved addition. They were found difficult by many children. The problems read:

5/6 In a game of darts Billy King had scored 187 and Jock Scott 223.

(a) What is the difference between these scores? (problem 5)

(b) How many more does Billy need to make 301? (problem 6)

9 Farmer Till had 210 sheep. At the market he sold 88 and bought 25. How many sheep has he now?

Phase 1. Here, out of thirty-nine children, eleven children managed to solve problem 5, five solved problem 6, and problem 9 was correctly solved by eight children. The types of difficulties experienced by the children unable to solve these problems were of a varied nature. They were evident in the area of comprehension, in transforming and in process skills, as indicated by the error analysis.

Phase 2. The rewritten versions used four main categories of difficulties for their creation. These were: comprehension difficulties, transforming difficulties, process skills difficulties and cognitive overload. The type of changes that were made for these categories were:

- for comprehension difficulties wording was altered or explained, graphics were removed if unnecessary, or layout was altered

- for transforming difficulties the numerical aspect of the problem was presented as a straightforward sum using the standard algorithm - this is referred to as a "sum only" problem below

- for process skills the need to carry was removed

- for cognitive overload the numbers were reduced to below sixty

These changes are discussed and explained in Chapter 6 pp. 174-183 and will not be further explored here.

The rewritten versions for problem 5 produced only one significantly easier version. This was the "sum only" version. However, the versions examining process skills, cognitive overload and one of the comprehension versions were successfully solved by more children than was the original. Thus there was a trend towards increased facility though this was not statistically significant.

For problem 6 this trend (that the rewritten versions were easier) was confirmed. Only one of the rewritten versions did not produce a statistically significant difference, this was one of the versions examining comprehension difficulties.

The rewritten versions to problem 9 were all significantly easier than the original. The straightforward addition "sum only" was particularly easy, followed by the cognitive overload version.

Phase 3. Here the original problem was presented alongside the rewritten versions to a different sample. The number of children managing to solve problem 5 was similar to Phase 1. However, statistically significant differences were found for three of the five rewritten versions. Only those examining comprehension difficulties were not significantly different from the original.

For problem 6 the number of children managing to solve the original was greater. Two out of the rewritten versions - "sum only" and cognitive overload - produced statistically significant differences, making these two rewritten versions easier. This coincides with problem 5. The rewritten version examining process skills was not significantly easier though the trend was in that direction.

Three rewritten versions were statistically significant for problem 9. The two "sum only" versions - one subtraction, one addition - and the cognitive overload. Interestingly enough the cognitive overload version was almost as easy as the addition sum.

These results were considered in detail in Chapters 5, 6 and 7. Thus only the overall implications of these findings will be considered here. There are a number of factors that seem to be in evidence here:

- the fact that the "cognitive overload" versions was significantly easier suggests that more children than those who responded successfully to the original problem can understand the language used in these type of problems. These cognitive overload versions retained the original language of the word problem from whence it stemmed. Despite the fact that they required the children to read the problem, transform, and solve it they were on the whole easier than the subtraction sums "sum only" versions stemming from the same problem and presented using the standard algorithm.

- the fact that the process skills version which removed the need to carry was significantly easier in problem 5 and 9, and that the trend was in the same direction for problem 6, suggests that particular aspects of the subtraction algorithm may be obscure to many children.

- These two factors - cognitive overload and process skills demanding carrying - examined in conjunction with the responses to the "sums only" versions examining subtraction skills show an interesting finding: in problem 9 there was a statistically significant difference between the subtraction sum and the cognitive overload problem ($\chi^2=7.02$, DF=1, $p<0.01$). The cognitive overload, despite the need for reading and transforming into a sum was the easier. This may support the suggestion that children use a different form of representation for these problems. That is the representation is different from that taught in the classroom based on the standard algorithm. This alternative representation is discussed below.

Representation of subtraction problems. It was suggested in Chapter 7 (p. 239) that children's informal methods for these types of problems do not always coincide with the standard method advocated in the classroom. It was suggested that one reason for the improved performance on the cognitive overload versions was that children could successfully solve these problems mentally as the smaller numbers used lessened the demand on the memory. It cannot be confirmed from the data gathered in Phase 3 that this was indeed the case. However, inspection of the responses show lack of workings for many of these versions. This could indicate support for the suggestion that informal methods were used.

Evidence in support of informal methods from Phase 1. It is interesting to note that examination of transcripts from Phase 1 shows that out of the eleven who successfully solved problem 5 three used the informal adding on method. One of these when asked to record what he had done (script 12) wrote the following:

```
 187
+???
 223
```

For problem 6, one of the five successful children used the adding on method. This may not seem a very large proportion. However, considering that the adding on method has not been part of the taught curriculum, and that the standard algorithm has received considerable attention, it is worth considering this informal method further. Examination of the scripts also show that several of the children who were unsuccessful attempted to solve these problems by adding on. So, for example, one child when trying to work out the answer to problem 6 was asked: "what are you trying to do - how are you trying to work it out?". His answer was: "trying to see how many more it goes up to 301". (script 38)

Thus the way the child chooses to represent the problem may be affected not only by the wording but also by the size of the numbers involved. It is also likely that it is affected by the interpretation the child puts on the situation. In certain situations they may think that they have to represent the problem "the way the teacher does it". This was possibly evident in some scripts. One child during Phase 1 (script 7) consistently represented the subtraction problems using the standard subtraction algorithm. However, she showed very little evidence of understanding this representation. None of these problems were solved correctly.

Establishing specific factors that consistently affect problem difficulty does not seem realistic at this stage. If the child interprets the problem-solving situation as requiring the standard algorithm, and this method is poorly understood, the problem will not be successfully completed. However, the same child may be successful using an alternative method that has developed from the child's own informal knowledge. Thus not only factors within the problem affects its likelihood of successful solution. The child's interpretation of the situation and the emphasis that the teacher puts upon the method is also of importance. It is felt therefore that to pursue internal word problem structure is probably not a useful way forward for research in this area.

8.3 An evaluation of approaches used within this project. This project used a combination of methods to access children's understanding of the word problems in their textbook. Three main aspects need to be considered here: (i) the effectiveness of the methods used; (ii) any alternative routes to understanding of children's word problem solution behaviour that could have been more effective than the one employed; and (iii) the limitations on this project in terms of generalisability.

8.3.1 Types of methods used. Both qualitative and quantitative methods were used in this project. The main ones were the task-based interview and the paper-and-pencil test. These two will be discussed separately and the implications of their use assessed.

The task-based interview. This type of method, which is virtually identical to the clinical interview, was pioneered by Piaget, and has been used by many researcher wanting to gain deeper understanding of children's cognitive development. Its development was discussed in Chapter 3 and this aspect will not be considered further here. What needs to considered are its strengths and weaknesses in relation to this project.

The greatest strength of the task-based interview that it allows the researcher to explore the child's understanding of a problem with the child as the task is carried out. Thus any solutions, or the lack of a solution, can be explored so that the actual difficulty can be more precisely assessed.

Its main weakness is that it is very time-consuming and it is therefore difficult to gather enough data for generalisations and predictions to be made. It depends on the sensitivity and expertise of the interviewer. It was felt that this research might have been improved had the researcher had a better initial training in using the task-based interviews with children.

It was felt that it was a useful tool for this project, and that valuable data was obtained that could not have been gathered in any other way. It could probably have been improved by the use of video-recording which would allow for deeper analysis of the overall situation. In this study only tape-recorded, transcribed records were analysed. However, the use of video-recording was outwith the scope of this study and would also have introduced its own limitations in terms of flexibility of use of time and space.

The paper-and-pencil tests. Again this method has been considered in Chapter 3. It has for long been the main method to examine competence in mathematics in general. Its strengths, weaknesses and overall usefulness is considered below.

The greatest advantage of paper-and-pencil tests is that they provide a quick and easy to score method of gathering large amounts of data. It is an effective way of determining problem difficulty in a large population (as used in Phase 2/3), or in giving an indication of children's mathematical ability (as used in selection for the sample in Phase 1).

The main weakness of this tool is that it normally only classifies data as correct or incorrect. This method therefore does not allow for the identification of the strategies used for the solution of the problem. When this method is used for testing children's understanding of mathematics it is not necessarily very accurate as it is difficult to create test items that effectively test children's understanding.

This method did provide an acceptable way of selecting a sample, and for comparing two samples. It also provided a means of assessing relative problem difficulty in Phase 2 and 3. It is felt, however, that had the paper-and-pencil tests been used in conjunction with selective task-based interviews in Phase 2 and 3, this would have yielded greater understanding of children's problem solving behaviour.

Choice of methods for the project. It was felt that the choice of methods was useful for a study of this nature but that, with hindsight, a number of improvements could be made. These relate mainly to more effective training in the use of the task-based interview, and further selective use of task-based interviews in concjunction with paper-and-pencil tests.

8.3.2 Alternative ways of investigating children's understanding of word problems. In this project 181 children were involved. Their understanding of a small number of word problems used for practising three arithmetic operations was studied. It provided interesting data but only a small number of "performances" of problem solving behaviour was collected from each child. Learning takes place over a much longer time span. It is possible that studies of this nature would be enriched by using smaller number of children in a case-study approach. This would involve sampling their performance on these types of problems not only over the period of six months but for two or three years. Changes in understanding of word problems could be more effectively monitored in this way. However, the case-study approach would introduce its own limitations in that generalisability would be further curtailed. The generalisability of this project is discussed below.

8.3.4 Generalisability of this project. The concept of generalisability is of importance in educational research if research findings are to be applied on the wider educational scene. This means that the sample used within the research must be as representative as possible of the wider population from which it was drawn, and that the tasks the children are asked to do are representative of the types of tasks they usually carry out. It is also essential that the situation the children find themselves in is similar enough to that of the classroom. If the latter is not the case the children may interpret the situation to be different from that of the classroom and behave differently thus influencing the results. In view of this it is possible to consider the generalisability of this study in terms of: (i) the sample; (ii) the tasks used; and (iii) the setting of the project.

(i) The sample. The main sample for Phase 1/2, which was studied most intensively, consisted of forty children of "average" ability from five different classes. The schools represented quite a broad range of different types of catchment areas but were all city schools. It is probably fair to say that the schools involved were reasonably representative of the types of schools to be found within a Scottish city. However, the sample was small and thus it was possibly atypical or skewed. The notion of what is "average" was discussed in chapter 5 and the suggestion was made that there was no such thing as the "average" pupil. Hence perhaps it is not possible to pursue the notion of the typical sample too far. Perhaps it is better to suggest that a larger sample is more likely to contain a more evenly spread variation than was found within this project.

(ii) The tasks. The tasks that were used were part of the children's normal "diet" of mathematics tasks in that they were taken from their standard textbook. Only a small number were used, and in that sense limitations are imposed. However, a project of this type is essentially of a limited nature and a greater number or diversity in problems would have been outwith the time limits of this study.

(iii) The setting. The research was school-based and thus closer to the normal school environment than laboratory based research. However, the actual interaction of one-to-one with the researcher did not represent the typical classroom setting. The expectations that the children have in such a setting as to what the task involves may well differ from those they have in the classroom and so affect their performance on the task. Thus to generalise to how they are likely to be able to solve word problems in the classroom setting is not valid. However, what was being investigated was children's understanding of the type of word problems they had encountered in the classroom setting. This setting did afford one way of investigating this understanding and would be as effective as the laboratory and more illuminating than simply studying the child in the classroom setting. Future problem solving behaviour in the classroom setting cannot be validly predicted, but a comment can be made on previous learning experiences in the classroom.

Thus a study of this nature is limited in its general applicability. However, it has afforded some interesting insights: some of these are in line with other research, and others, possibly because of the difference in the tasks used, differ from other research. These latter findings point towards the complexity of this type of research. This type of investigation should form part of a larger study in order to gain greater generalisability as the research of De Corte and Verschaffel (1989) suggest.

8.4 An evaluation of the concepts used to investigate children's problem solving behaviour.

Three main concepts were used to interpret the children's word problem solving behaviour: (i) schema, (ii) informal strategy and (iii) cognitive overload. The first two of these stemmed from previous research, the third was researcher created, but drawing on cognitive research into reading. Two other areas also require careful examination: (iv) the use of performance to investigate children's learning; and (v) the use of the word difficulties when explaining some of the problem solving behaviours. These five areas will be investigated below.

8.4.1 Schema. Schema theory was discussed in length in Chapter 2 and this discussion will not be repeated here. It has been a widely used theory mainly within language learning but also, perhaps more latterly, within mathematics. It is an intuitively plausible theory. It does, as suggested in Chapter 6, provide a means for explaining diversity in behaviour. However, there is a danger that it may have become nothing more than a convenient coathanger - that is, it is used to explain behaviour that cannot be explained in any other terms. It is a wide-ranging theory in that it tries to explain all human behaviour, yet it cannot be tested empirically and thus it cannot, to use Popper's ideas, be falsified. It may therefore not be useful to explain research that ultimately aims to have practical application. An attempt has been made by Riley et al (1983) amongst others to use schema theory to explain the development of mathematical understanding (see Chapter 2 pp. 36-39). These theories develop the notion of schemata specifically developed to deal with mathematical understanding. Riley et al has been criticised by Carpenter for using only limited data when developing these schemata - that which fits in with the theory. Not considered by Carpenter but which perhaps also should be considered is that this type of theorizing seems to see mathematical understanding as developing in isolation from other types of understanding. For many people their most effective use of mathematics is possibly in an intuitive manner in situations where it has a practical application. For example, a child may quite effectively work out the amount of change s/he would have left after making a purchase. Give the same child the same calculation in a textbook and s/he may be quite lost. Thus what might be termed "a sense of number" can be developed independently of the formal mathematics taught at school. This was particularly well illustrated by one of the children that participated in this study (script 15).

This perhaps illuminates the difficulties in using schema theory to explain mathematical behaviour. For research purposes it may be more useful to employ contrasting concepts that allow exploration of certain aspects of education. In relation to mathematics understanding the concepts of conceptual and procedural knowledge have already been explored within this study. Others that suggest themselves are: teaching and learning, action and reflection, and formal and informal methods. These would operate at different levels of education, some being more overarching concepts such as teaching and learning, whilst formal and informal methods would be subsidiary concepts. Not only would these concepts be studied in isolation but their interrelationship would be explored. Within teaching and learning may be found considerable tensions. The action/reflection concepts may also suggest tension - teaching requires constant action and response from the teacher. Action from the pupil is considered helpful to learning. Yet for learning to occur reflection is required by both the teacher and the learner. Models drawing from different organisational levels could perhaps be created from these types of investigations. They would reflect the individual - pupil or teacher; the group - classes or groups of teachers; or the institution; or several institutions. Not only would relationships within the different levels be examined but also those between different levels. These would maybe provide more useful insights for educationists than does schema theory.

8.4.2 Informal strategy. This concept has been used by many researchers, e.g. Carpenter & Moser, to indicate the means that children use for obtaining a solution without using standard procedures such as those taught in schools. In this project the main alternative to the standard, taught algorithms was found within subtraction where some children used an adding-on rather than subtractive procedure. The term informal strategy is perhaps not the most effective to explain these problem-solving behaviours. In a sense the term procedure is perhaps more appropriate. Strategy tends to imply an overall plan of action - in this project in fact what was being considered was a procedure that was employed to achieve the goal. Thus procedure is subsidiary to strategy and would be a more effectively descriptive term.

The use of "informal" is also problematic as it immediately brings into consideration the notion of "formal". In some sense formal may be interpreted as being more correct than informal. That is not the interpretation intended in this project. What is being considered is the procedure that is considered most effective by the education system and therefore specifically taught in schools. Informal methods tend to have grown from the child's own understanding of the problem. Hence informal methods have the advantage of being "anchored" in the child's own understanding. Against this must be considered that formal mathematical methods have been developed over centuries and are not likely to grow naturally without any formal tuition. The distinction made by Van Lehn (op. cit.) between natural and non-natural learning and discussed in Chapter 2 is perhaps of relevance here. Two aspects need to be considered: firstly, if informal procedures can be employed in the classroom setting thus building on the child's previous knowledge and also perhaps enhancing the understanding of more formal procedures; and secondly if all the formal procedures need to be taught to all children when calculators are available. Thus whether the the word informal should be used to describe a child's procedure could be debated. However, it does seem still to provide a means of differentiating learning that stems from the formally taught school-based learning to that which grows out of the overall environment that the child experiences. Provided neither is considered inferior they can usefully be employed for investigating how children learn. The implications of their relative effectiveness in future learning can also be considered.

8.4.3 Cognitive overload. This term was created by the researcher to try to explain why children seemed to be able to cope with similar word problems in certain circumstances but not in others. An example of this was reducing the numbers in problems 5, 6 and 7. It was suggested that a certain area in the brain acted as a "clearing house" when problem solving occurred and that this clearing house was of limited capacity. The term provided a useful vehicle for exploring these ideas. However, it must be considered that this is only a plausible suggestion and the difficulties may equally well be caused by lack of knowledge or an inability to locate or retrieve the necessary information. Thus the term may well have outlived its usefulness with the end of this project. The "cognitive overload" category was only *introduced in Phase 2* which relied on paper-and-pencil tests for gathering data and the limitations of this type of data gathering was discussed in Section 8.4. It was suggested that further use of task-based interviews during this stage may have increased the understanding of children's word problem solving behaviour.

8.4.4 The use of performance to investigate learning. This project set out to investigate children's understanding of word problems and to examine the types of difficulties they experience. Ultimately this was done in order to understand how children learn. However, only a small number of instances of problem solving performances were examined. Learning is a slow and long-term progress and instances of performance do not necessarily illuminate the path of learning. This project does have a usefulness though in that it illustrates that very recently learnt concepts are not necessarily well retained by many children after a short period.

8.4.5 The use of the term difficulties when describing some problem solving behaviours. When a child was unable to reach a solution s/he was considered to have a difficulty in a particular area. It may be that this difficulty is transient, and that with maturity and further teaching it will disappear. In that sense it is not so much a difficulty as a passing phase that perhaps has to occur for learning to happen. The link with the previous section on using performance to investigate learning becomes apparent here: if only instances of performance are used to investigate learning what is really a passing phase may erroneously become termed a difficulty. Thus, in many educational settings, a correct answer tends to lead to the inference that learning has taken place, an incorrect answer assumes lack of learning. The discussion quoting Silver (Chapter 2, p. 42) is also of relevance here. Silver suggested that evidence for conceptual understanding often relied on the child showing what might be only procedural knowledge. Only research of a longitudinal nature could attempt to sort out this conundrum. Other research literature and evidence from the secondary school would suggest that at least some of the difficulties experienced by the children in this project will remain as difficulties, at least in the school setting.

8.5 A consideration of theoretical and practical implications of this project.

Theoretical implications. The implications for theory have already been considered to some extent in the discussion on the use of schema theory. Perhaps the most important function of a theory is to provide a framework within which research can be carried out. It helps formulate research questions and provides suitable concepts to examine the findings. As far as this project is concerned it has been found useful to draw on some aspects of schema theory to examine the children's behaviour. However, on reflection it is felt that further research in this area may proceed more usefully if it employs a number of contrasting concepts related to education and also examines the interrelationship and tensions between these concepts. These types of concepts were explored in Section 8.4.1 when discussing the usefulness of the concept of schema.

Practical implications. This project set out to investigate children's understanding of word problems as these problems appeared in their textbook. It has provided useful insights into how children attempt to solve these problems. It has not provided a "recipe" for how word problems should be taught. However, two aspects should be considered before concluding this project: (i) the "reality" aspect of word problems, and (ii) suggestions from other research on the use of word problems in teaching mathematics and their relationship to the findings of this project.

(1) The ability of word problems to provide a "realistic" context. The intention of using word problems in the classroom is to provide a realistic setting for practising mathematics. Chapter 1 included a brief discussion quoting Bell and Stephens (pp 7-9) as questioning the reality of many of the word problems used in the average textbook. The criticisms voiced by these researchers are valid. Many of the settings of the word problems used in this project provided little reality for the children who were asked to solve them. It seems, on reflection that it is extremely difficult for any textbook to provide a selection of word problems that will have validity for large numbers of different children. Even within the small main sample of forty children in this project the home backgrounds were distinctly varied and thus the experiences that the children related these problems to varied. Perhaps a more effective way of teaching word problems would be to encourage teachers to make up word problems that would relate more realistically to the type of background that the children came from. Within the typical primary school there are a number of events, some of them occurring daily, that could perhaps be used within the mathematics curriculum. One such example is the daily collection of dinner money. Here a variety of mathematical tasks could be created ranging from simple ones of adding up the daily amounts to those requiring subtraction and multiplication. Weekly, monthly and term totals could be worked. For older children averages could be introduced - on average how many school dinners are served per week or per month. The list is endless and will not be pursued further. Again Van Lehn's discussion of natural and non-natural learning springs to mind. Schools seem to be very effective at turning what could be a natural learning situation into a non-natural learning situation by creating, what sometimes must be, unnecessary barriers between the actual task and its occurrence in daily life outside the school.

(ii) Suggestions from other research for the teaching of word problems. Chapter 2 (pp. 53 ff.) considered two contrasting approaches within word problem research. Nesher & Teubal, Carpenter & Moser and De Corte & Verschaffel amongst others base their research on semantic *structures whilst Lean* et al suggest that psycholinguistic theory provides better explanations of children's problem solving behaviour. These two approaches offer to some extent contrasting approaches to the teaching of word problems. Lean et al suggests that children's *understanding* of language develops more slowly than generally expected and therefore their understanding of mathematical terms will be affected. Thus great care needs to be taken when introducing children to a variety of word problems. Carpenter, De Corte & Verschaffel argue, on the other hand, that children's understanding of the semantics is limited because their diet of word problems is limited to only certain types of structure. They advocate much greater variety. This project would perhaps offer some cautious support for Lean et al in that many children appeared to have a poor understanding of some aspects of the mathematical use of certain phrases -

in particular those relating to subtraction. However, it was found that when the problems using this language were changed so that the numbers involved were considerably reduced, more children could understand the structure. Perhaps the advice of Carpenter that word problems should not be used to practise newly learnt arithmetic routines but as an exploration of mathematical relationships is valid. If this was the case, the children would presumably be free to create their own representation of the problem and this could then lead to an exploration of alternative ways of reaching a solution. The standard algorithm could then be taught following this type of exploration. Word problems would thus become a step in the progress towards attaining understanding of the standard algorithm and not a means for practising routines already learnt in abstraction from a real life setting.

To sum up, this research intended to examine children's understanding of word problems. In order to do so the intended role of word problems also needs consideration. If they are to play a useful part in increasing children's mathematical understanding it could be argued that they should be more realistically representative of the child's immediate environment, and that they should afford an opportunity for an exploration of mathematical relationships evident within that environment.

8.6 Conclusion.

This chapter has given an overview of the whole project in terms of its structure and its findings. It has been suggested that valuable insights have been gained into children's word problem solving behaviour. However, in terms of generalisability the project is limited due to the small sample and the fact that only a small number of problem solving performances were used to investigate learning. The tools used within the project were useful but improved training in interviewer technique and more flexible use of the task-based interview method would probably have enhanced the project. The concepts used to explain the children's behaviour provided a useful theoretical framework. However, for any future projects of this nature it may be that different conceptualization may provide more fruitful research findings. The role of word problems in mathematics teaching has been discussed. The form in which these appear in the textbook cause them to be seen as unnatural routine exercises rather than providing a link with the natural environment as intended by the textbook writers.

BIBLIOGRAPHY

ANDERSON, R.C. (1984) Some reflections on the acquisition of knowledge. *Educational Researcher.* **13** (9), 5-10.

BADDELEY, A.D. (1976) *The psychology of memory.* New York, Basic Books.

BAROODY, A.J. & GINSBURG, H.P. (1986) The relationship between intial meaningful and mechanical knowledge of arithmetic. In Hiebert, J. (ed.) *Conceptual and procedural knowledge: the case of mathematics.* Hillsdale, NJ. Lawrence Erlbaum Associates.

BELL, A.W., COSTELLO, J. & KUCHEMANN, D. (1983) *A review of research in mathematical education. Part A. Research on learning and teaching.* Windsor, NFER-Nelson.

BELL, F.H. (1980) Posing and solving verbal problems. *Mathematics Teacher.* **December,** 625-656.

BERNSTEIN, B. (1979) Social class, language and socialization. In Lee, V. (ed.) *Language development.* London, Croom Helm in association with the Open University Press.

BRANSFORD, J.D. & McCARRELL, N.S. (1977) A sketch of a cognitive approach to comprehension: some thoughts about understanding what it means to comprehend. In Johnson-Laird, P.N. & Wason, P.C., (eds.) *Thinking: readings in cognitive science.* Cambridge, Cambridge University Press.

BRITTON, B.K., GLYNN, S.M. & SMITH, J.W. (1985) Cognitive demands of processing expository text: a cognitive workbench model. In Britton, B.K. & Black, J.B. (eds.) *Understanding expository text.* New Jersey, Lawrence Erlbaum Assoc.

BRYANT, P. (1975) The understanding of invariance by very young children. In Whitehead, J.M. (ed.) *Personality and learning 1.* London, Hodder & Stoughton, in association with the Open University Press.

BYERS, V. & ERLWANGER, S. (1985) Memory in Mathematical understanding. *Educational Studies in Mathematics,* **16,** 3, 259-281.

CAMPBELL, J.I.D. (1987) The role of associative interference in learning and retrieving arithmetic facts. In Sloboda, J.A. & Rogers, D. (eds.) *Cognitive processes in mathematics.* Oxford, Clarendon Press.

CARPENTER, T.P. (1986) Conceptual knowledge as a foundation for procedural knowledge: implications for research on the initial learning of arithmetic. In Hiebert, J. (ed.) *Conceptual and procedural knowledge: the case of mathematics.* Hillsdale, NJ. Lawrence Erlbaum Associates.

CARPENTER, T.P., HIEBERT, J. & MOSER, J.M. (1981) Problem structure and first-grade children's initial solution processes for simple addition and subtraction problems. *Journal for research in mathematics education,* **12**, 1, pp. 27-39.

CARPENTER, T.P., MOSER, J.M. & BEBOUT, H.C. (1988) Representation of addition and subtraction word problems. *Journal for Research in Mathematical Education.* **19** (4) 345-357.

CLEMENTS, M.A. (Ken) (1980) Analyzing children's errors on written mathematical tasks. *Educational Studies in Mathematics.* **11**, 1, 1-21.

Collins Concise Dictionary of the English Language. (1978) London, Collins.

DAVIS, R.B. (1984) *Learning mathematics: the cognitive science approach to mathematics education.* London, Croom Helm.

DE CORTE, E. & VERSCHAFFEL, L. (1989) Teaching word problems in the primary school: what research has to say to the teacher. In Greer, B. & Mulhearn, G. (eds.) *New directions in mathematics education.* London, Routledge.

DE CORTE, E. & VERSCHAFFEL, L. (1987) Using retelling data to study young children's word-problem-solving. In Sloboda, J.A. & Rogers, D. (eds.) *Cognitive processes in mathematics.* Oxford, Clarendon Press.

DEPARTMENT OF EDUCATION AND SCIENCE (1982) *Mathematics counts. Report of the committee of inquiry into the teaching of mathematics in schools under the chairmanship of Dr. W.H. Cockcroft.* London, H.M.S.O.

DESFORGES, C. & COCKBURN, A. (1987) *Understanding the mathematics teacher. A study of practice in first schools.* Lewes, Falmer Press.

Developing Mathematical Thinking, EM235 (1982), Milton Keynes, Open University.

DONALDSON, M. (1978) *Children's minds.* Glasgow, Fontana/Collins.

Edinburgh Reading Tests, Stage 1. (1977). Edinburgh, Edinburgh University.

GELMAN, R. GALLISTEL, C.R. (1978) *The child's understanding of number.* Cambridge, M.A., Harvard University Press.

GELMAN, R. & MECK, E. (1986) The notion of principle: the case of counting. In Hiebert, J. (ed.) *Conceptual and procedural knowledge: the case of mathematics.* New Jersey, Lawrence Erlbaum.

GINSBURG, H.P. & ALLARDICE, B.S. (1983) Children's psychological difficulties in mathematics. In Ginsburg, H.P. *The development of mathematical thinking*. Orlando, Academic Press Inc.

GRAHAM, D.J. (1987) An associative retrieval model of arithmetic memory: how children learn to multiply. In Sloboda, J.A. & Rogers, D. (eds.) *Cognitive processes in mathematics.* Oxford, Clarendon Press.

GREER, B. & MULHERN, G. (1989) Improving mathematics education: a human problem. In Greer, B. & Mulhearn, G. (eds.) *New directions in mathematics education*. London, Routledge.

GROEN, G. & KIERAN, C. (1983) The many faces of Piaget. In Ginsburg, H.P. (ed.) *The development of mathematical thinking.* Orlando, Academic Press Inc.

HEINEMANN EDUCATIONAL (1989) *Primary and middle school catalogue.* Oxford, Heinemann.

HENDERSON, D. (1991) Static classes raise research eyebrows. In *Times Educational Supplement Scotland.* 12.4.91. 5.

HIEBERT, J. (1990) The role of routine procedures in the development of mathematical competence. In Cooney, T.J., (ed.) *Teaching and learning mathematics in the 1990's.* National Council of Teachers of Mathematics.

HIEBERT, J. & LEFEVRE, P. (1986) Conceptual and procedural knowledge in mathematics: an introductory analysis. In Hiebert, J. (ed.) *Conceptual and procedural knowledge: the case of mathematics.* Hillsdale, NJ. Lawrence Erlbaum Associates.

JERMAN, M. (1973) Problem length as a structural variable in verbal arithmetic problems. *Educational Studies in Mathematics.* **5** (2) 109-123.

JERMAN, M. & REES, R. (1972) Predicting the relative difficulty of verbal arithmetic problems. *Educational Studies in Mathematics.* **4** (3) 306-323.

JOHNSON-LAIRD, P.N. & WASON, P.C., eds. (1977) Introduction to inference and comprehension. In Johnson-Laird, P.N. & Wason, P.C. (eds.) (1977) *Thinking: readings in cognitive science.* Cambridge, Cambridge University Press.

KANE, R.B. (1967) The readability of mathematical English. *Journal of Research in Science Teaching.* **5**, 296-298.

KANE, R.B. (1970) The readability of mathematics textbooks revisited. *Mathematics Teacher.* **6**, 579-581.

LABOV, W. (1979) The logic of nonstandard English. In Lee, V. (ed.) *Language development.* London, Croom Helm in association with the Open University Press.

LANGFORD, P.E. (1986) Arithmetical word problems: thinking in the head versus thinking on the table. *Educational Studies in Mathematics,* **17**, 2, 193-199.

LEAN, G.A., CLEMENTS, M.A.(KEN) & DEL CAMPO, G. (1990) Linguistic and pedagogical factors affecting children's understanding of arithmetic word problems: a comparative study. *Educational Studies in Mathematics,* **21**, 165-191.

LINVILLE, W.J. (1976) Syntax, vocabulary, and the verbal arithmetic problem. *School, Science and Mathematics.* 152-158.

McINTOSH, A. (1981) When will they ever learn? In Floyd, A. (ed.) *Developing mathematical thinking.* Wokingham, Addison-Wesley in assoc. with Open University Press.

MARSHALL, S.P. (1983) Sex differences in mathematical errors: an analysis of distractor choices. *Journal for Research in Mathematics Education.* **14**, 4, 325-336.

Mathematics Across the Curriculum, Unit 1, Introduction (PME233) (1980). Milton Keynes, Open University Press.

MINSKY, M. (1977) Frame-system theory. In Johnson-Laird, P.N. & Wason, P.C. (eds.) *Thinking: readings in cognitive science.* Cambridge, Cambridge University Press.

MORRIS, P. (1974) *Methodology and two early approaches. Block 1, Unit 2 (DS261).* Milton Keynes, Open University.

NESHER, P., GREENO, G. & RILEY, S. (1982) The development of semantic categories for addition and subtraction. *Educational Studies in Mathematics,* **13**, 4

NESHER, P. (1976) Three determinants of difficulty in verbal arithmetic problems. *Educational Studies in Mathematics,* **7**, 369-388.

NESHER, P. & KATRIEL, (1977) A semantic analysis of addition and subtraction word problems in arithmetic. *Educational Studies in Mathematics,* **8**, 251-269.

NESHER, P. & TEUBAL, (1975) Verbal cues as an interfering factor in verbal problem solving. *Educational Studies in Mathematics.* **6** (1) 41-51.

NEWELL, (1977) On the analysis of human problem solving protocols. In Johnson-Laird, P.N. & Wason, P.C. (eds.) (1977) *Thinking: readings in cognitive science.* Cambridge, Cambridge University Press.

NICHOLSON, A.R. (1977) Mathematics and language. *Maths in School.* 32-34.

OATLEY, K. (1978) *Computational metaphors for perception. Block 2, Unit 8 Cognitive Psychology (D303).* Milton Keynes, Open University.

ORTON, A. (1987) *Learning mathematics: issues, theory and classroom practice.* London, Cassell.

RESNICK, L.B. (1983) A developmental theory of number understanding. In Ginsburg, H.P. *The development of mathematical thinking.* Orlando, Academic Press Inc.

RESNICK L.B. & FORD, W.W. (1984) *The psychology of mathematics for instruction.* N.J. Lawrence Erlbaum Associates.

RIDGWAY, J. (1987) *A review of mathematics test.* Windsor, NFER-Nelson.

RILEY, M.S., GREENO, J.G. & HELLER, J.I. (1983) Development of children's problem-solving ability in arithmetic. In Ginsburg, H.P. *The development of mathematical thinking.* Orlando, Academic Press Inc.

SCHANK, R.C. & ABELSON, R.P. (1977) Scripts, plans and knowledge. In Johnson-Laird, P.N. & Wason, P.C. (eds.) *Thinking: readings in cognitive science.* Cambridge, Cambridge University Press.

SCOTTISH PRIMARY MATHEMATICS GROUP (1985) *Primary mathematics. A development through activity. stage 2, 2nd ed. Teacher's Notes.* London, Heinemann.

SCOTTISH PRIMARY MATHEMATICS GROUP (1985) *Primary mathematics. A development through activity. stage 2, 2nd ed. Textbook.* London, Heinemann.

SHUARD, & ROTHERY, (1984) *Children reading mathematics.* London, John Murray.

SIEGLER, R.S. (1987) Strategy choices in subtraction. In Sloboda, J.A. & Rogers, D. (eds.) *Cognitive processes in mathematics.* Oxford, Clarendon Press.

SILVER, E.A. (1986) Using conceptual and procedural knowledge: a focus on relationships. In Hiebert, J. (ed.) *Conceptual and procedural knowledge: the case of mathematics.* Hillsdale, NJ. Lawrence Erlbaum Associates.

SKINNER, B.F. (1975) The science of learning and the art of teaching. In Whitehead, J.M. (ed.) *Personality and learning, 1.* London, Hodder & Stoughton, in association with the Open University Press.

SLACK, J. (1978) *Semantic memory. D303, Block 3, Units 18-19.* Milton Keynes, Open University.

SMEDSLUND, J. (1979) Piaget's psychology in practice. In Floyd, A. (ed.) *Cognitive development in the school years.* London, Croom Helm, in association with the Open University Press.

STEPHENS, M. (1977) Mathematics: medium and message. *Maths in School.* **6**, 2-5.

STILL, A. (1983) Psychology and mathematics. In Nicholson, J. & Foss, B. (eds.) *Psychology Survey No. 4.* Leicester, British Psychological Society.

VAN LEHN, K. (1986) Arithmetic procedures are induced from examples. In Hiebert, J. (ed.) *Conceptual and procedural knowledge: the case of mathematics.* New Jersey, Lawrence Erlbaum.

VAN LEHN, K. (1983) On the representation of procedures in repair theory. In Ginsburg, H.P. (ed.) *The development of mathematical thinking.* Orlando, Academic Press Inc.

VON GLASERFELD, E. (1987) Learning as a constructive activity. In Janvier, C. (ed.) *Problems of representation in the teaching and learning of mathematics.* New Jersey, Lawrence Erlbaum Assoc.

WATSON, (1980) Investigating errors of beginning mathematicians. *Educational Studies in Mathematics,* **11**, 3, 319-329.

YOUNG, D. (1979) *Y mathematics series - Y1. Sevenoaks, Hodder & S*toughton.

APPENDIX A

Problems used for the Exploratory study (Chapter 4) and Phase 1 (Chapter 5) of the Main study.

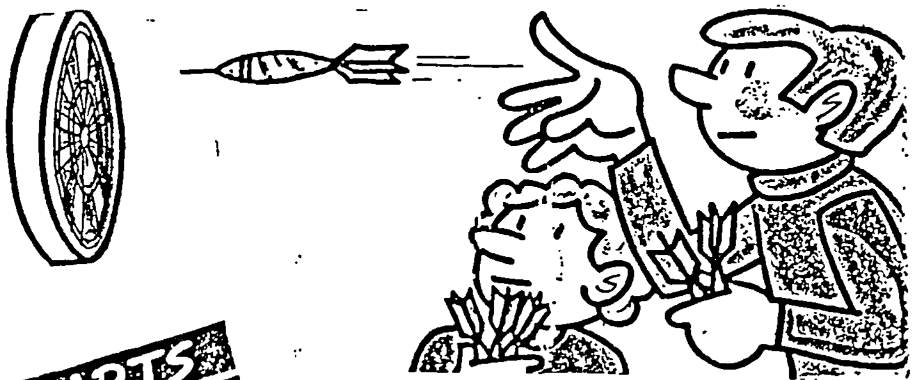
1 Here are the marks given to a skater by the judges.
Find the total mark.

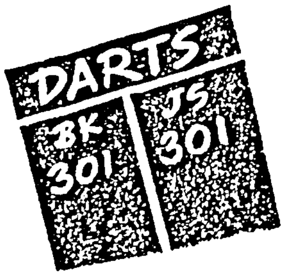
2 Jim enters the 80 metres race and is given a start of 13 metres.
How far does he have to run?

3 One morning 37 boys and 46 girls go to the library.
That afternoon 39 boys and 59 girls go.

How many (a) boys, (b) girls go to the library that day?

4 Altogether the boys take out 67 books and the girls 87 books.
How many books are taken out that day?

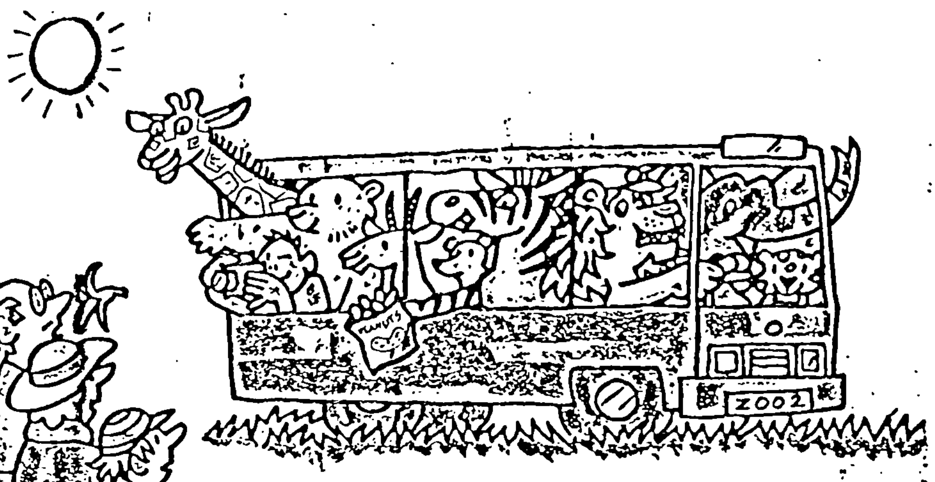




During a game of darts Billy King had scored 187 and Jock Scott 223.

5 (a) What is the difference between these scores?

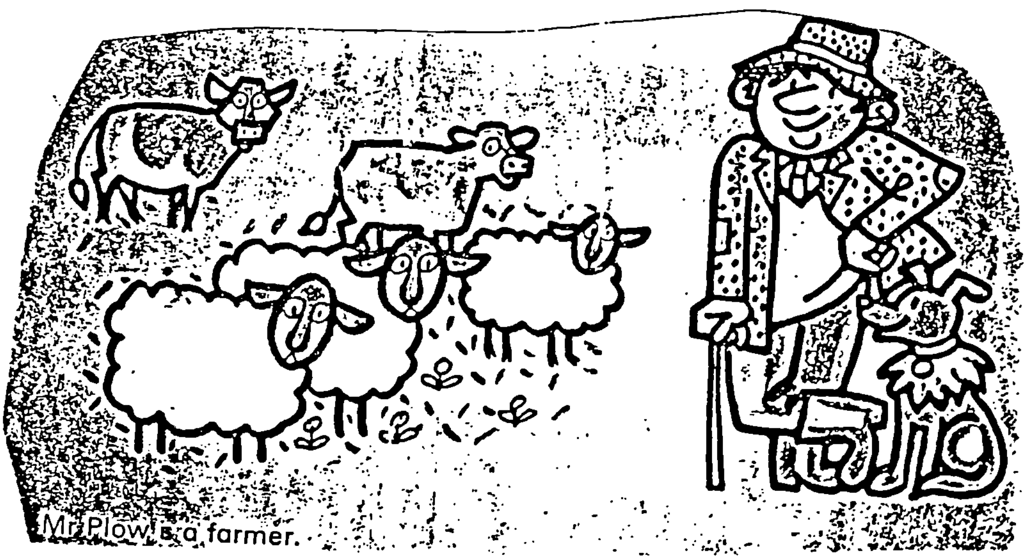
6 (b) How many more does Billy need to make 301?



This is a wild-life safari park.

7 (a) There are 745 zebras and 813 antelopes. How many more antelopes are there than zebras?

8 (c) Last year there were 247 giraffes. This year their number has increased to 300. By how many has their number increased?

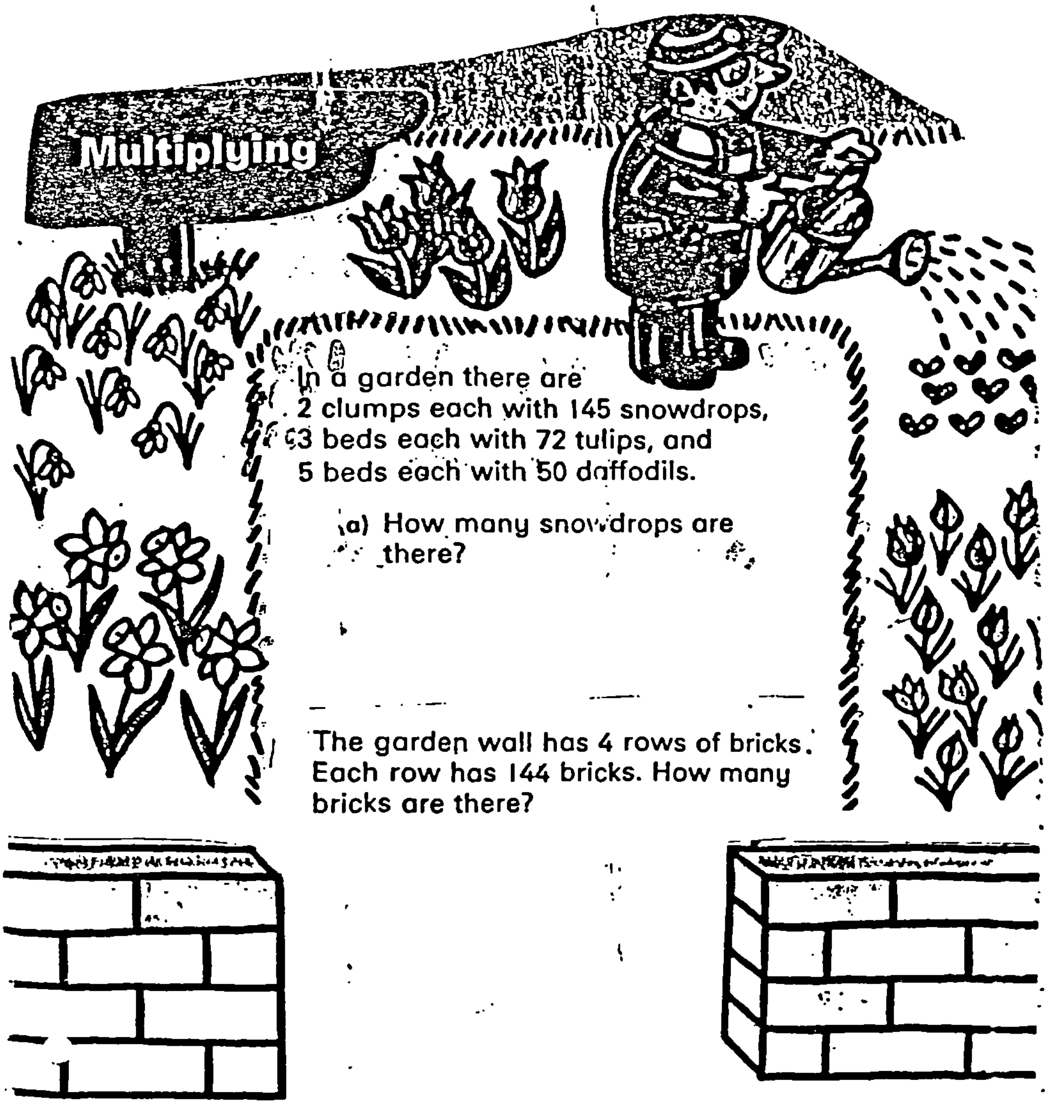


Mr Plow is a farmer.

9 (d) Farmer Till had 210 sheep. At the market he sold 88 and bought 25. How many sheep has he now?

10 (a) There are 4 classes in Park School.
Each class has 32 pupils.
How many pupils is this altogether?

11 (b) Each class gets a box of 48 pencils.
How many pencils is this altogether?

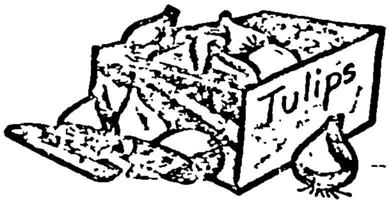


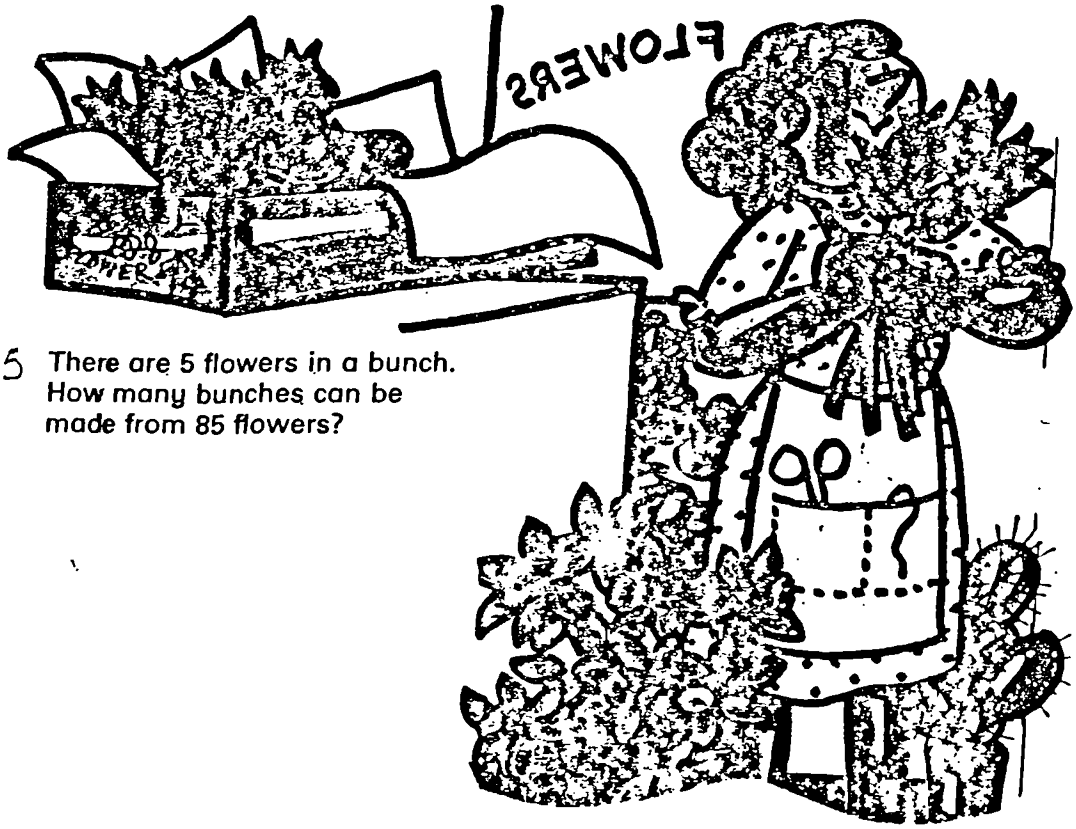
12 In a garden there are
2 clumps each with 145 snowdrops,
3 beds each with 72 tulips, and
5 beds each with 50 daffodils.

(a) How many snowdrops are there?

13 The garden wall has 4 rows of bricks. Each row has 144 bricks. How many bricks are there?

14 A gardener has a box of 65 tulip bulbs.
He plants them in 5 bowls, with the same number in each bowl.
How many bulbs does he plant in a bowl?





15 There are 5 flowers in a bunch.
How many bunches can be made from 85 flowers?

16 A box holds 77 squares. How many shapes like this can be made?

APPENDIX B

MATHEMATICS TEST
FOR PRIMARY 4

- *Please put your name and your age on the front of this booklet.*
- *Read the questions carefully.*
- *Put your answer in the space beside or underneath the question.*

NAME: ____

AGE: ____

Addition

1. 25 + 34

2. Add

```
 674
+496
```

3. June's ribbon is 20 centimetres long, Mary's 33 centimetres and Jane's 29 centimetres. What is the total length of the ribbon?

4. In a game of darts Jim scored 124, 96 and 147. What was his total score?

Subtraction

1. Subtract 17 - 4

2. Subtract

```
 65
-41
```

3. Subtract

```
 539
-278
```

4. Write the missing numbers in place of the dots.

```
 947
-..6
 57.
```

5. Jane has 287 British stamps in her album. If she has 902 stamps altogether, how many of them are not British?

Multiplication

1. Multiply 3×7

2. Multiply 8×6

3. Multiply
$$\begin{array}{r} 245 \\ \times\ \underline{\ \ 5} \end{array}$$

4. Multiply
$$\begin{array}{r} 48 \\ \times\ \underline{7} \end{array}$$

5. 29 children in a class were each given six sweets. How many sweets were given out altogether?

6. Multiply
$$\begin{array}{r} 204 \\ \times\ \underline{20} \end{array}$$

Division

1. Divide $35 \div 5$

2. Divide $4\overline{)85}$

   Answer ___
   Remainder ___

3. 64 oranges are divided equally into 8 bags. How many oranges are in each bag?

4. Divide $7\overline{)112}$

5. Divide $8\overline{)469}$

   Answer ___
   Remainder ___

Relationships

1. Here are four numbers. Look at them carefully. Which is the biggest? Draw a ring round it.

3472 2473 7324 4273

2. Two numbers are missing from this series. Write them down in the spaces.

4, 8, 12, ..., 20, ..., 28, 32

3. The numbers in the top row go with the numbers in the bottom row in a certain way. Fill in the missing numbers.

18	8	16	30	(b)
9	4	(a)	15	14

4. Write the number nine thousand and nine in figures.

5. Another way of writing twenty-five tens is

25 250 25000 2570

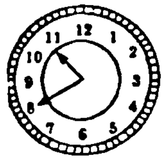
Underline your answer.

Time

1. If today is Monday what is the day after tomorrow?

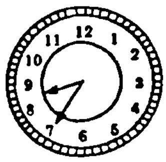
2. What is the time on the clock? Ring your answer.

A. Twenty minutes past eleven.
B. Eight minutes to eleven.
C. Twenty minutes to eleven.
D. Five minutes to eight.
E. Five minutes past eight.



3. This clock is 15 minutes slow. What is the correct time? Ring your answer

A. 8.50
B. 8.35
C. 8.20
D. 8.15
E. 7.20



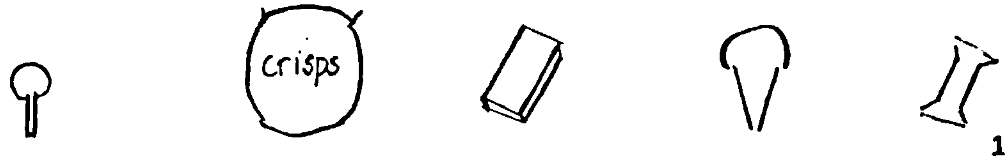
4. How many hours are there in 2½ days?

5. In the month of June there were 13 sunny days, 9 cloudy days and the rest were rainy days. How many rainy days were there?

Money

1. What change would I get from 10p if I buy a bag of crisps?

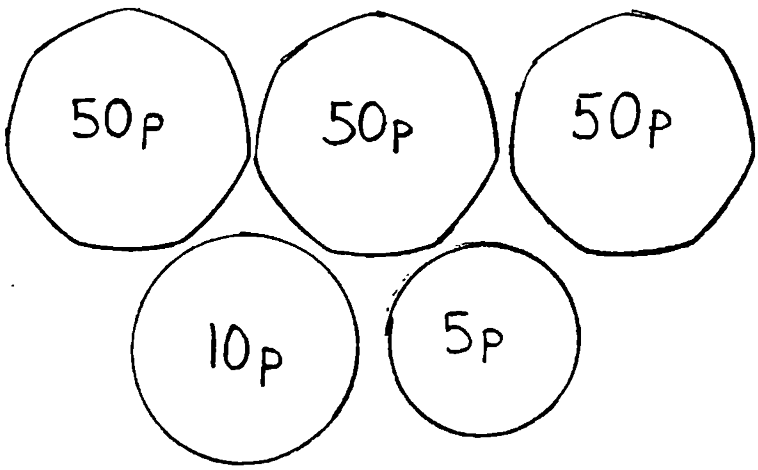
______p



lollipop 2p crisps 7p biscuit 5p cone 3p chew 2p

2. You go shopping with a 50p piece. You buy two tins of soup each costing 20p. How much change will you bring home?

3. What is the total value of these coins?



4. Ian wants to buy a ball. He has 68p. He needs 27p more. What is the price of the ball?

5. A chocolate bar costs 13p. How much will 6 of these bars cost?

6. I have seven 5p coins in my purse. I buy a ball costing 16p. How much money have I left?

APPENDIX C

The following transcript is included to show how the transcripts were scored according to the error analysis. Chapter 5 discussed the concept of the "average" pupil and suggested that within this group there was considerable diversity. This transcript has therefore not been chosen to represent all the "average" pupils - rather it shows some difficulties, such as subtraction, that is common to many of the children, and others that are idiosyncratic to this pupil.

The following classification of difficulties was used:

1. Reading: (i) word recognition
   (ii) symbol recognition
   (iii) graphics
2. Comprehension: (i) general understanding
   (ii) specific terms
   (iii) identifying relevant information
3. Identification of operation
4. Transforming
5. Process skills (i) faulty computation
   (ii) random response
   (iii) no response
   (iv) careless slip
6. Encoding

Only the number with the relevant subcategory will be indicated on the script.

TRANSCRIPT OF INTERVIEW WITH AIMEE – 24.11.89

Problem 1

Int: Could you do No 3 for me? Can you find No 3?

Aimee: Here are the marks given to a skater by the judges. Find the total marks ....

Int: What do you do with that?

Aimee: Well, you add them up.

Int: Mm, you can either do it on paper or in your head or whatever.

Aimee: You write them down and then you add them up, so you put an add sign there. ..... 8 add 8 is 16, and another 8 is 24 and a 6 .... 30 ... another 7, 37 and a 9 .. 46

Int: Good

Aimee: So it's 46 altogether.

Int: That's right, it's 46 marks altogether. You knew to add these separately, didn't you. Have you seen a skating competition ever, when they get marks ... have you seen that ...?

Aimee: Yeah.

Problem 2

Int: Could you do No 6 for me.

Aimee: Jim enters the 80 metres race and is given a start of 13 metres. How far does he have to run? You go .... 13 to 80 .. so you add. 4

Int: You add, do you? What do you think?

Aimee: You add. 4

Int: You add. What do you think "is given a start of" means?

Aimee: 80 .... 13 and it would be 93 .... } 2(i)

Int: Mm, if that's what it is. If he's given a start of 13 do you think he has to run the whole race? ..... You don't think so? If you have a race that starts there and finishes there and that's all 80 metres. You don't think he has to run the whole bit. Maybe he starts a bit in, doesn't he? He doesn't have to run 13 metres. How would you find that out? Cause this bit .. would that be more than 80 or less than 80? (this conversation is accompanied by a drawing and Aimee's responses are very faint)

Aimee: Less

Int: Yes, so how could you work it out that way?

Aimee: Take away ...

Int: You can write it as a sum, write it down.

Aimee: You can't go 0 take away 3 you score that one out ... so it's 10 take away ... 7 ..... so it's 67 ..

Int: Yes, so that's how far he has to run, isn't it? And what is it ....?

Aimee: miles 6

Int: Is it miles he's running?

Aimee: metres.

Int: Yes, so you write metres beside it. If it was miles it would be an awful long way ... like from here to Aviemore or something .... could you do 4a?

Problem 3

Aimee: How many (a) boys (b) girls go to the library that day? .... so one morning 37 boys and 46 girls go to the library. That afternoon 39 boys and 59 girls go. Why does it say 46?

Int: What do you mean? Why does it give the girls?

Aimee: Oh, the afternoon.

Int: Yes.

Aimee: How many ...

Int: Which one are you working out?

Aimee: Well, you're adding so you're going to go 37 .. 46 2 (iii

Int: What does it say .. how many ..

Aimee: (a) boy

Int: That's like a separate question, you know how to do this, you do those sums separately, don't you (referring to sums (a) - (d) at top of page), so you've got (a) and (b) is the second question, so what are you working out? ..

Aimee: girls 2 (iii)

Int: What is it? What's that word ... how many.. ?

Aimee: boys

Int: boys went to the library that day.

Aimee: 7 add 9 ... 16 .. you put down 6 .. carry .. 3 add 3 is 6 add 1 ...

Int: So what's your answer?

Aimee: 76

Int: 76 what?

Aimee: boys.

Int: Good. You don't need to do (b). Just write boys beside it .. could you do No 6 as well.

Problem 4

Aimee: Altogether the boys take 67 books and the girls 87 books. How many books are taken out that day? ... (mumbles) .... 7 and 7 is 14 .. 8 and 6 is 14 add 1 15.

Int: That's right, cause 8 and 6 is the same as 7 add 7, if you took 1 from the 8 and gave it to the 6 you'd have 7 add 7, wouldn't you? A quick way of working it out, isn't it? What is it?

Aimee: 154 books.

Int: Good. Could you do 1a .. 1a.

Problem 5

Aimee: What is the difference between these scores?

Int: Mm, what does it say after 1?

Aimee: During a game of darts Billy King had scored ... oh, I know ... Is it add or take away? 2 (i)

Int: I thought you knew this one.

Aimee: I think we had it for homework.

Int: Mm, what do you think?

Aimee: Add. 3

Int: What are you trying to do, what are you trying to find out?

Aimee: The difference .... it's add. 3

Int: Would that give you the difference?

Aimee: No.

Int: No?

Aimee: (has written 187 4
-223)
7 take 3 is 4 ... 8 take 2 is 6 ... 1 take away 2 ... you can't .... 11 take away 2 is .. (adds 1 to make 187 into 1187) .....

Int: Where did you get that 1 from (added digit)? Mm? where did you get it from? Did you score anything out to get it? Just plucked it out of nowhere, didn't you! Can you do that 2 take 1? ... No, so what have you done here? What have you done? You know what you have done done with the two sums (meaning numbers)?

Aimee: MM.

Int: How do you have to write them when you do a take away?

Aimee: The bottom one up there and the other one down there.

Int: You write the bigger number on top, don't you? Is that right? Let's see you do it then.

Aimee: 3 take away 7 is 3 .. 5(i)

Int: Can you do it?

Aimee: so you score out .... mumbles ..

Int: So, what's the difference between the scores?

Aimee: 36

Int: Good, can you do (b) as well.

Problem 6

Aimee: How many more does Billy need to make 301? .... 2(i)

Int: How do you think you can work that one out? .. Don't know? How many more does he need to make 301? What sort of a sum do you think you've got to do?

Aimee: Add 3

Int: Add, do you think so?

Aimee: Take away ..

Int: You think so?

Aimee: Add 3

Int: One of them is right.

Aimee: Multiply. 3

Int: Multiply, no. Let's have a think about it. Look at this (draws diagram - two hoops). I've got 7 sweeties (in one hoop) and 4 sweeties in there (other hoop).

this one?

Aimee: 3

Int: 3 more, don't I. Like that, what sort of sum could I do, if I added them, that's 4 in that one and 4 in that one. Do I get the right answer?

Aimee: It's a take away.

Int: If I take what do I get ... 3 don't I, which is what I had to add. So to find out how many more ....

Aimee: Multiply. 3

Int: No, that's right, it's take away ... That's what you said. To find out how many more .. I did that sum, didn't I? I took that lot from that lot, didn't I? And that told me how many more I needed to add to it.

Aimee: So it's a take away.

Int: It's a take away. It's (b) you're doing ...

Aimee: Do you do the same as you did there ..? 4

Int: Which way do you have to set them up? Which number do you have to have on top?

Aimee: The smallest.

Int: The smallest one on top?

Aimee: The biggest.

Int: The biggest, cause otherwise you don't have enough to take away from, do you?

Aimee: take away ... 187 ..

5(i)

Int: Can you take 1 from there (0)? .. No, so where do you have to go? That's it .. when you do that, you have to put something here, what does the 0 become...? It's 10 and then you have to cross it out again and then what does it become?

Aimee: 9

Int: right ..

Aimee: mumbles ...

Int: So what .. how many more does he need?

Aimee: 114

Int: Good, will you remember howw to that next time, do you think?

Aimee: Was I right at the sum I did on Friday?

Aimee: Was I right at the sum I did on Friday?

Int: When you did the test? How did you write them then?

Aimee: Did I do them wrong?

Int: I can't remember. I'll have a look and see what you did. .. Is that what you've made mistakes with? (looks at maths test) Yes, you've got .... the take away one ..No, you've got that one right, are there any more take away? You haven't got that one right. 43-3 .. I bet you could that ...

Aimee: 40.

Int: Yes, it was just not thinking straight .. and that one, I bet you could do as well. 7-4 ... it's 3 isn't it .. I think you were in a rush, doing it.

Aimee: Yes.

Int: Yes, that's what you have to remember that .. maybe you did make that mistake. So remember that next time, cause if you don't put the bigger number on top you don't have enough to take away from. Right, we're going to try this one .. 1a. Could you do 1a.

Problem 7

Aimee: There are 7 ... 745 and 813 anelopes. 1(i)

Int: Antelopes. Do you know that word. No, it's an animal a bit similar to that one. 2(ii)

Aimee: What is it?

Int: It's a deer almost.

Aimee: I think I've seen one but I'm not sure.

Int: Mm. I think you only get them in the zoo. You don't get them in the forest or anything else. Just in the zoo. I think they live mainly in Africa.

Aimee: Right. It's a take away again .....

Int: Good, and how did you know to do a take away that time?

Aimee: The bigger number on the top ...

Int: Mm, but why did you do a take away and not an add?

Aimee: Because it was .... if you had zebras and you wanted to find out .. if you were adding you would just get the wrong answer.

Int: Mm, you'd get far too many wouldn't you? It's another one where you're finding the difference, isn't it? You've got two groups and you're wanting to find out how many more there is in one or how many less in the

the other. Could you do 1c as well.

Problem 8

Aimee: Will you be letting us hear that as well (taperecorder)

Int: Mm. Do yo want to hear it.. Well, you do this and then I'll wind it back a bit and you can hear a little.

Aimee: Last year there were 247 giraffes. This year their number has increased to 200. By how many has their number increased? What does increased mean? 2(ii)

Int: It means to get more.

Aimee: Right 247 ...

Int: Do you think that's right ... what do you think it is?

Aimee: Is it take away? 3

Int: Do you think so? ... I think you're right. I think again you're looking at how many there were last year which is one group, and then how many there are this year. And you're wanting to find out how many more there were this year. So ..

Aimee: That would be right if it was adding?

Int: It would be right, you've added it right, but it's not the right sum for what they ask .. It says how many more has their .. increased ... the other way they could say is how many more are there now.

Aimee; This is difficult.

Int: It's a difficult one.

Aimee: 54

Int: 10 take 7 is .. 5(iv)

Aimee: 3

Int: Good, so it's 53 more .. what is it more of?

Aimee: giraffes

Int: Right, can you write giraffes. Could you do 1d for me?

Problem 9

Aimee: Last year ..

Int: 1d can you find 1d?

Aimee: Farmer Till had 210 sheep at the market he sold 88 and bought 25. How many sheep has he now. Take away

...

Int: That's right, but what else do you have to do.

Aimee: Add.

Int: Can you show me how you do that?

Aimee: Take away 8 ..

4

Int: Make sure you write the sum carefully so that the 8's are in the right place.

Aimee: Is that the wrong place?

Int: No, I think you're just about right. Let's see how you get on ... can you ..? Now think of it like ... (draws HTU diagram) .. hundreds, tens and units. Can you write them in the right place? Where do you write 210 ... ? Mm. 2 and 1 ten and a 0 .... that's right, now what about the 88? .. How many tens and how many units. Right . ...

Aimee: mumbles solution

Int: And what have you got left there?

Aimee: 1

5(i)

Int: Mm. You've got 0, you should have take away sign there, you've got 0 to take away, so you've got the 1 still, so how many sheep does have when he's sold 88?

Aimee: 122 .. now for the adding.

Int: Now for the easy adding.

Aimee; adds ...

Int: And where do you write the 25? Good ... so how many sheep does he have in the end?

Aimee: 147.

Int: Mm, sheep, could you write sheep in there, could you squeeze it in? Is this quite exhausting having to do all this maths? Just got a few more to do. Could you do 5a.

Problem 10

Aimee: 5a ... There are 4 classes in Park School. Each class has 32 pupils. How many pupils does the school have altogether? We've had this one again.

Int: Have you, so what do you do there?

Aimee: 4 x

Int: Good.

Aimee: 4 2's are 8 ...

Int: So what have you got .. 128 what?

Aimee: people. 1(i)

Int: Is it people, what does the word say? .. Pupils, do you know what a pupil is?

Aimee: Yes.

Int: What is it?

Aimee: Mm, teacher and pupil ... children.

Int: Yes, you're a pupil, aren't you?

Aimee: Yes.

Int: Good.

Aimee: s

Int: Yes, it would have to be cause it's more than one .. could you do b as well, 5b.

Problem 11

Aimee: Each class gets a box of 48 pencils. How many pencils is this altogether? ......

Int: How are you going to work that one out ..? Do you know how many classes there are?

Aimee: 7 2(iii)

Int: Are there 7? In that school? (Aimee's school has seven)

Aimee: Ah, 4 x 48

Int: Do you think so?

Aimee: 8 .... 4 .... 36 ... no 32 ..

Int: It's 192 what ..?

Aimee: pencils

Int: What does it say? Can you read it to me?

Aimee: pencils.

Int: Oh, what does it say at the top of that page?

Aimee: Multiplying by 5.

Int: by .. this one?

Aimee: Multiplying by 4

Int: Right, does that give you a clue, what you have to do?

Aimee: Yes, multiply by 4.

Int: That's what you've just been doing, haven't you. Do you read the bit at the top sometimes?

Aimee: Sometimes because it's sometimes on the board and it says the number of the page and what we're doing.

Int: Right, right. Sometimes it tells you what you're doing, but not always. Could you do 1a on this page?

Problem 12

Aimee: How many snowdrops are there? Multiplying .. is that them? (looks at illustration) 1(iii)

Int: Do you think those are snowdrops? What about reading that bit?

Aimee: In the garden there (misses out 2) are clumps each with 1 45 snowdrops, 3 beds each with 72 tulips, and 5 beds each with 50 daffodils. Snowdrops? (looks at illustration again) 1(iii)

Int: Those are snowdrops, yes I think ..

Aimee: 1, 2, 3, 4, ... 1(iii)

Int: What about reading this bit. How many snowdrops does it say there?

Aimee: 145 2(iii)

Int: And what does it say at the beginning here?

Aimee: 2 clumps ... 2 times ..

Int: Yes, what is a clump, do you know?

Aimee: No. 2(ii)

Int: No. It's a bunch of flowers growing in the ground, you know, sort of bunched together. So there are 2 of those bunches and each ... has how many ....

Aimee: 145 ... mumbles ... 2 ... 5(iv)

Int: Is it 3, you've done 2 x 4 and added that 1 (carried), so you've already added that 1 ... so what's 2 x 1? So how many snowdrops are there?

Aimee: 290 ..

Int: Mm, snowdrops ... can you write snowdrops beside it... what have you written there? Good. Could you do No 2 as well.

Problem 13

Aimee: The garden wall has 4 rows of bricks. Each row has 144 bricks. How many bricks are there?

Int: So what do you think you have to work out in that one?

Aimee: Mm, 2 times? 2 (iii)

Int: Read it again.

Aimee: 4 times.

Int: because there are 4 rows .... in fact they show you 1, 2, 3, 4 .. but there are 144 in each of those ones .. so.

Aimee: 4 4's are 16 ....

Int: Is that right? Good!

(Impulsive and untidy, but willing to discuss, listen and contribute)

Aimee's workings

1

8
9
8
6
7
+ 8
46

2

7
8̶10
−13
67 metres

3

3

37 / 46

37
+39
76 boys

~~46~~

4

67
+87
154 books

187
−223
964

5

1 2 13
~~2~~ ~~2~~ 13
−187
036

5

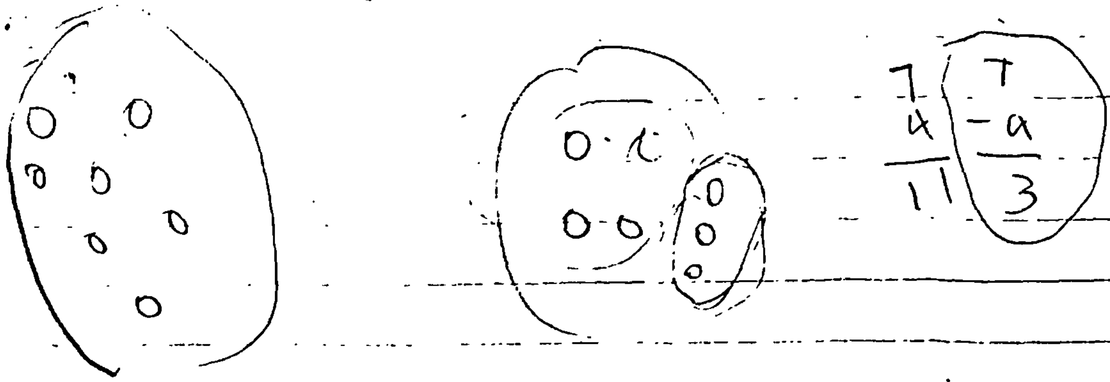
6

3 9
~~3~~ ~~10~~ 11
−1 8 7
1 1 4

7

7 10
~~8~~ ~~1~~ 13
−7 4 5
0 6 8

Researcher's diagram Problem 6



Second attempt

9

H	T	U
~~2~~ 1	~~1~~ 10	10
−	8	8
1	2	2

H	T	U
1	2	2
+	2	5
1	4	7 sh.

First attempt

8

~~247~~
~~+300~~
~~547~~

First attempt 29

18 ~~310~~ 10

247
05~~4~~3 giraffes.

First attempt

9

~~210~~
~~−88~~
~~0~~

10

32
x 4
128 pupils

11 48
x34
192 pencils

12 145
x 2
290 snowdrops

13 144
x 4
576 bricks

APPENDIX D

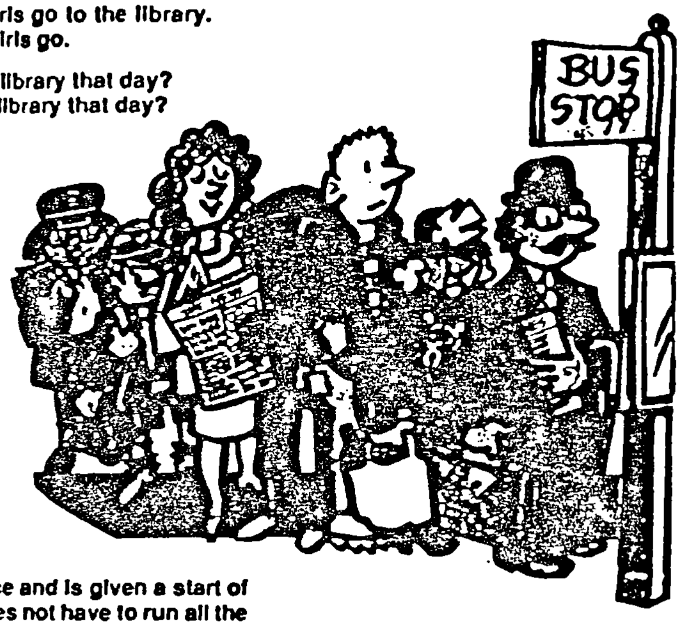
Problems used of Phase 2 of the Main Study (Chapter 6).

Version A

Add or subtract

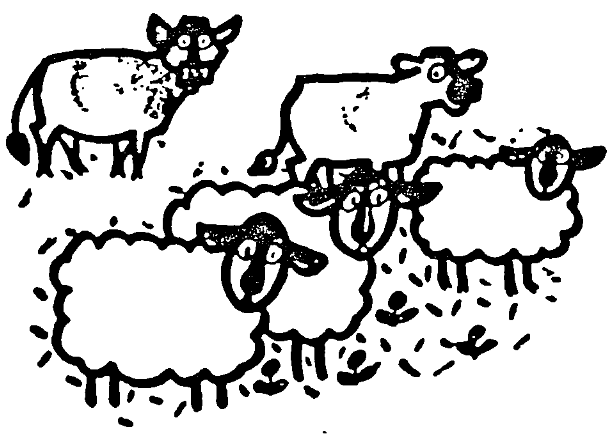
One morning 37 boys and 46 girls go to the library.
That afternoon 39 boys and 59 girls go.

1 a) How many boys go to the library that day?
b) How many girls go to the library that day?

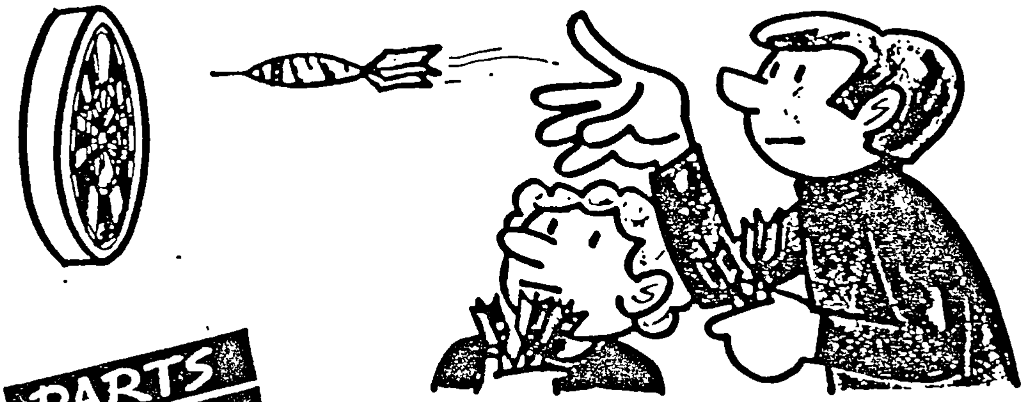


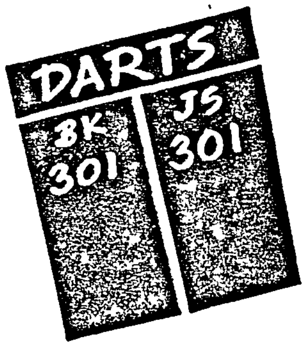
2 Jim enters the 80 metres race and is given a start of 13 metres. This means he does not have to run all the 80 metres. How far does he have to run?

3 a) Farmer Till had 210 sheep. At the market he sold 88 sheep. How many sheep has he now?
b) He then bought 25 sheep. How many sheep has he now?



Subtract or multiply

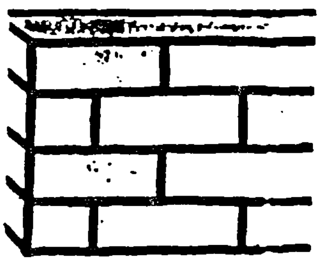




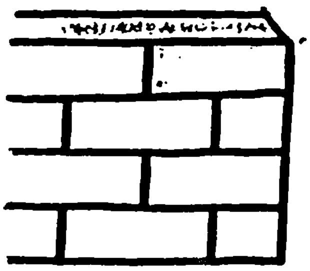
4 During a game of darts Bob Smith had scored 187 and David Brown 223.
a) How much less does Bob Smith have than David Brown?
b) What is the difference between Bob's score and a score of 301?

In a garden there are
2 clumps each with 145 primroses;
3 beds each with 72 lupins; and
5 beds each with 50 popples;

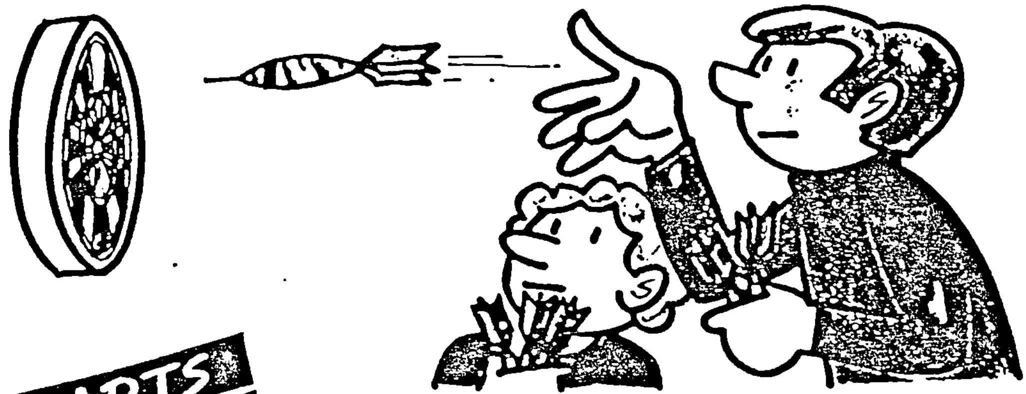
5 a) how many primroses are there?
b) how many lupins are there?
c) how many popples are there?

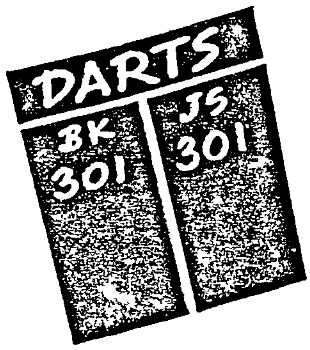


6 263
-197



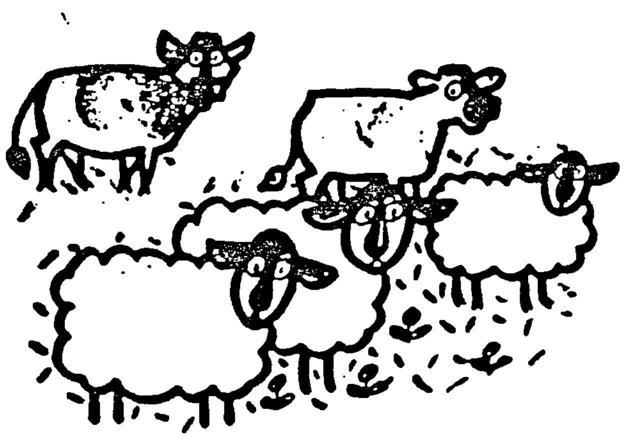
Add or subtract





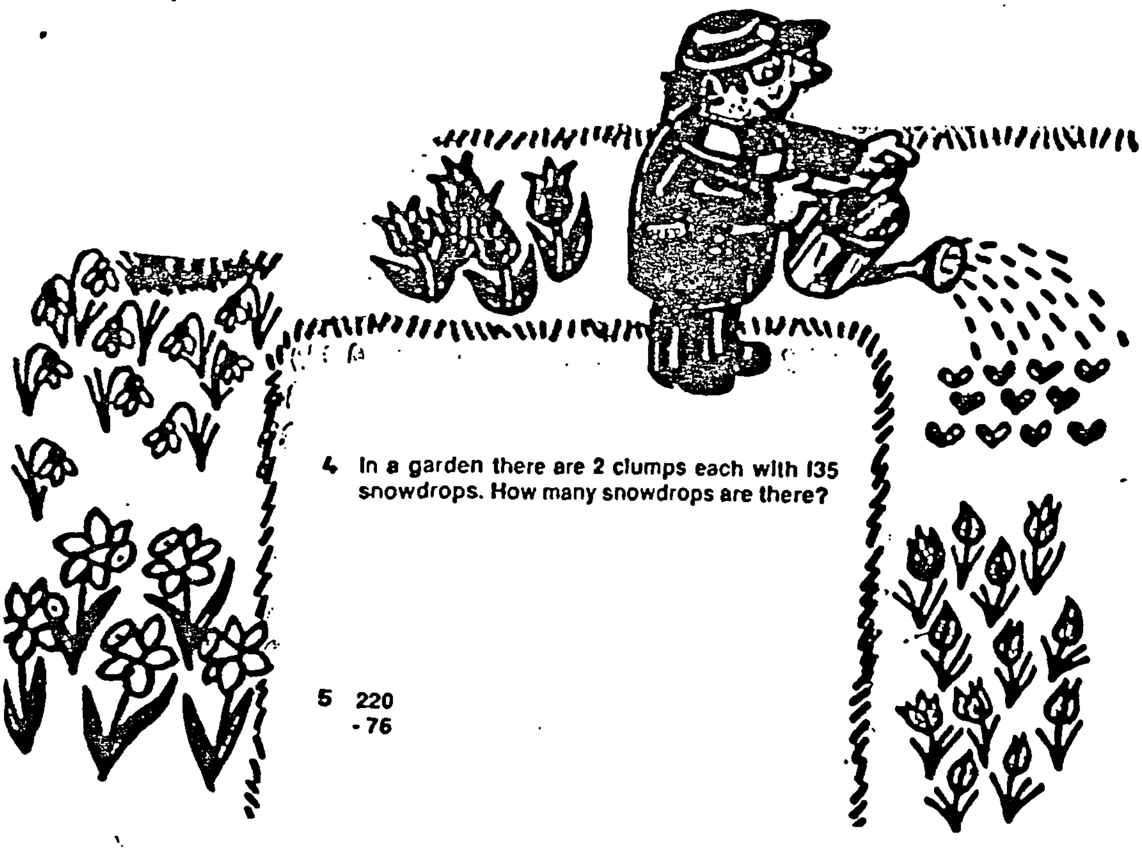
1 During a game of darts Colin White had scored 167 and Neil Stewart 213.
a) How much more does Neil Stewart have than Colin White?
b) Colin wants to make 303. How many less than 303 does he have?

2 Farmer Brown had 198 cows. At the market he sold 86 and bought 33. How many cows has he now?



3 Alison enters the 90 metres race and starts 12 metres in front of the others. How far does she have to run?

Multiply or subtract

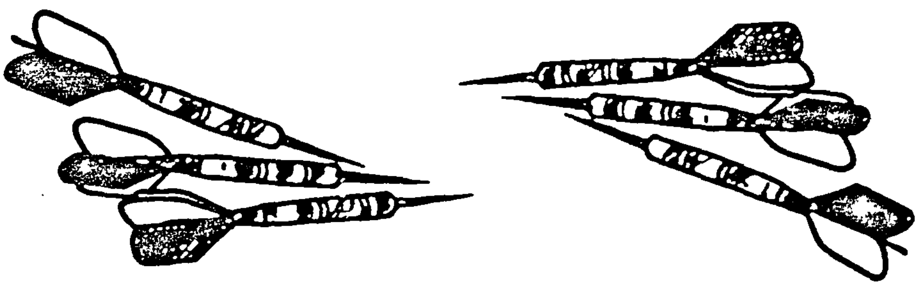


4 In a garden there are 2 clumps each with 135 snowdrops. How many snowdrops are there?

5 220
-76

6 During a game of darts Mike Wood had scored 36 and Jack Macdonald 43.

a) What is the difference between these scores?
b) How many more does Mike need to make 59?

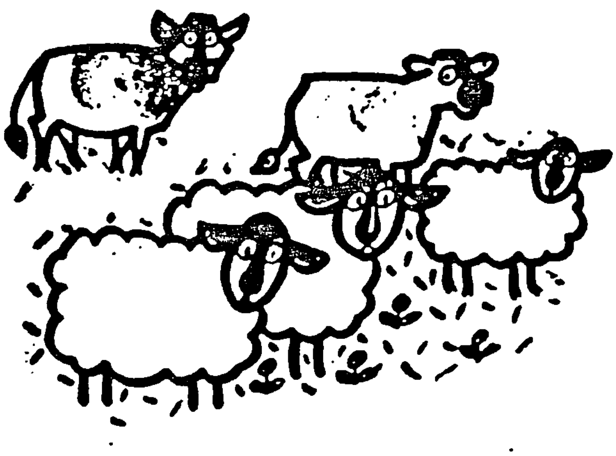


Add or subtract

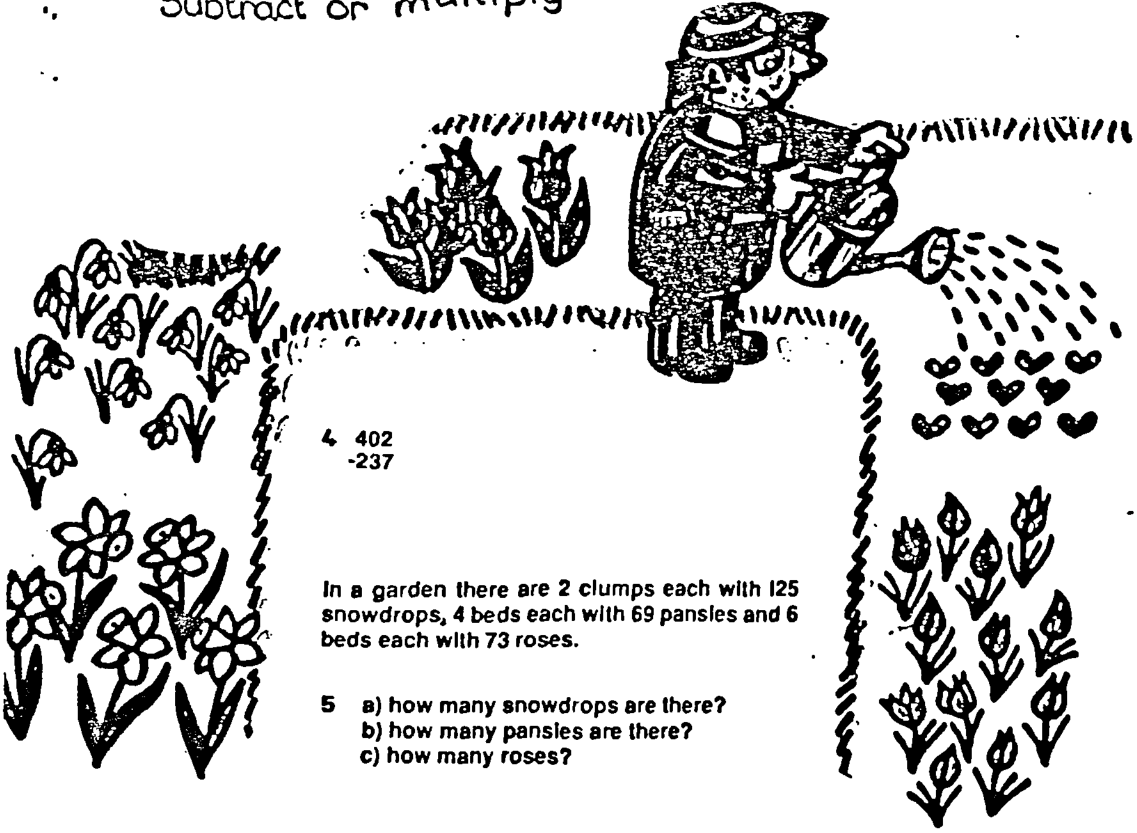
1 114
+ 33

2 David enters the 85 metres race and is given a start of 13 metres. How far does he have to run?

3 Farmer Macdonald had 60 sheep. At the market he sold 5 and bought 3. How many sheep has he now?



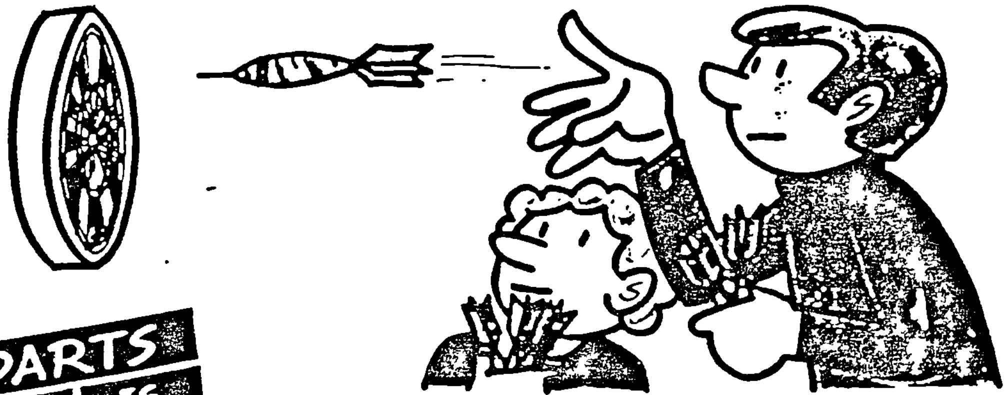
Subtract or multiply

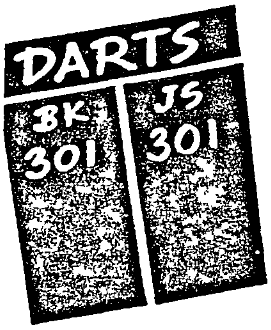


4 402
 -237

In a garden there are 2 clumps each with 125 snowdrops, 4 beds each with 69 pansies and 6 beds each with 73 roses.

5 a) how many snowdrops are there?
 b) how many pansies are there?
 c) how many roses?





6 During a game of darts Chris Smith had scored 123 and Bill Brown 235.
 a) What is the difference between these scores?
 b) How many more does Chris need to make 255?

APPENDIX E

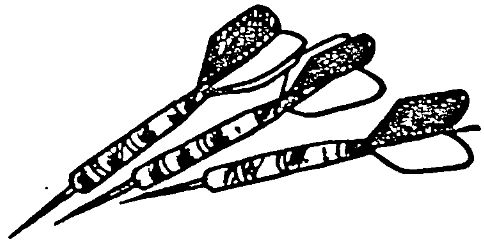
Problems used for Phase 3 of the Main Study (Chapter 7).

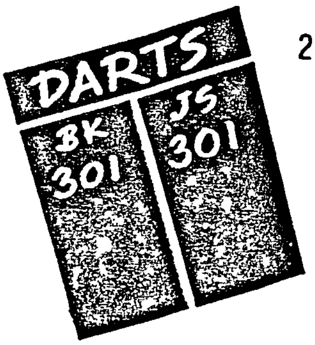
Version A

Add or Subtract

One morning 37 boys and 49 girls go to the library.
That afternoon 39 boys and 59 girls go.

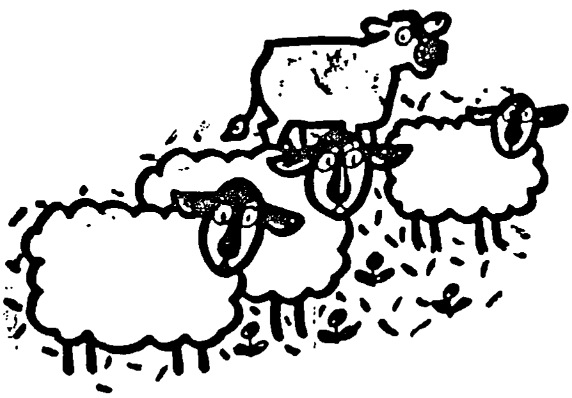
1 (a) How many boys go to the library that day?
(b) How many girls go to the library that day?



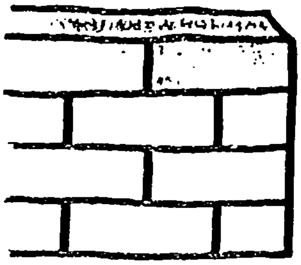


2 During a game of darts Billy King had scored 187 and Jock Scott 223.
(a) What is the difference between these scores?
(b) How many more does Billy need to make 301?

3 (a) Farmer Till had 210 sheep. At the market he sold 88 sheep. How many sheep has he now?
(b) He then bought 25 sheep, How many sheep has he now?

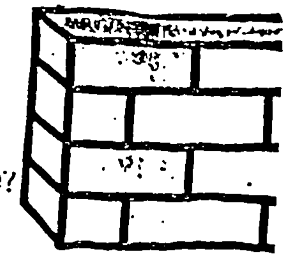


Subtract or multiply



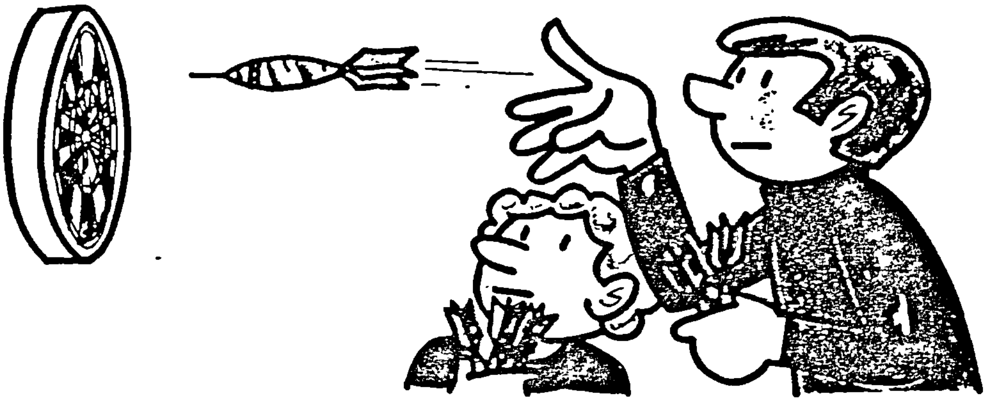
In a garden there are
2 clumps each with 145 primroses,
3 beds each with 72 lupins, and
5 beds each with 50 poppies.

4 (a) how many primroses are there?
(b) how many lupins are there?
(c) how many poppies are there?



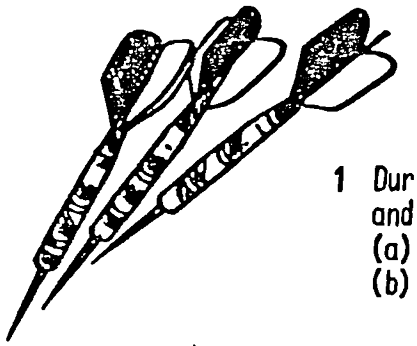
5 Jim enters the 80 metres race and is given a start of 13 metres. This means he does not have to run all the 80 metres. How far does he have to run?

6 263
-197



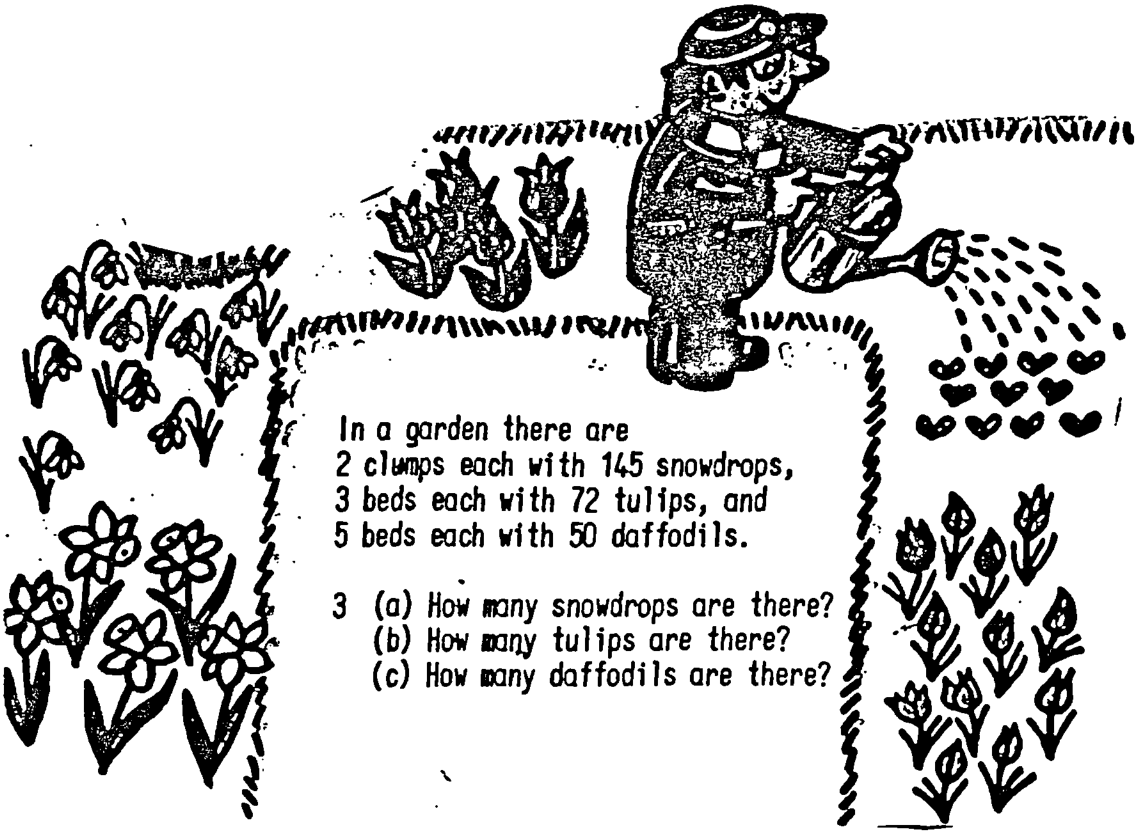
7 During a game of darts Bob Smith had scored 159 and David Brown 231.
(a) How much less does Bob Smith have than David Brown?
(b) What is the difference between Bob's score and a score of 301?

Add, Subtract or Multiply



1 During a game of darts Colin White had scored 167 and Neil Stewart 213.
   (a) How much more does Neil have than Colin?
   (b) Colin wants to make 303. How many less than 303 does he have?

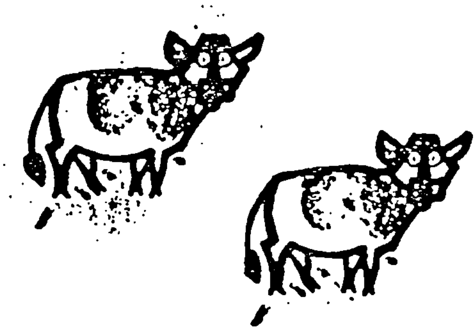
2 Alison enters the 90 metres race and starts 12 metres in front of the others. How far does she have to run?



In a garden there are
2 clumps each with 145 snowdrops,
3 beds each with 72 tulips, and
5 beds each with 50 daffodils.

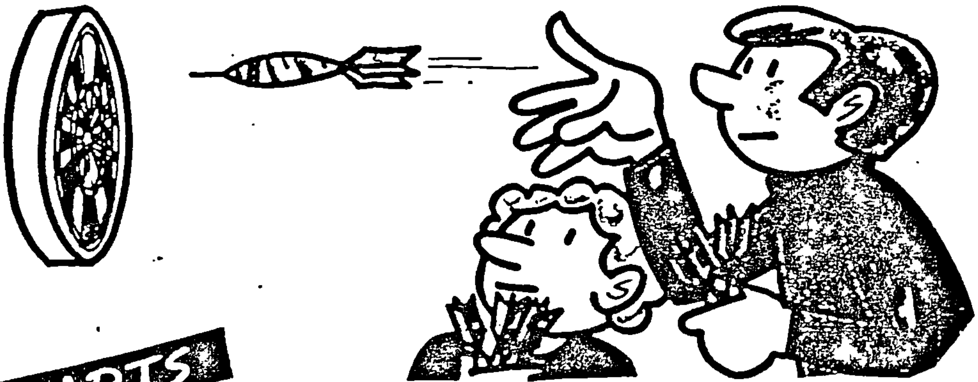
3 (a) How many snowdrops are there?
   (b) How many tulips are there?
   (c) How many daffodils are there?

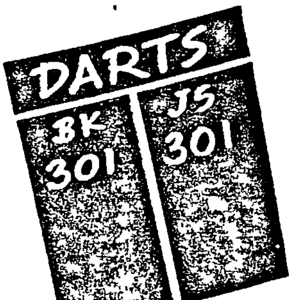
Add or Subtract



4 Farmer Brown had 198 cows. At the market he sold 86 and bought 33. How many cows has he now?

5 $\begin{array}{r} 220 \\ -\ 76 \\ \hline \end{array}$





6 During a game of darts Mike Wood had scored 36 and Jack McDonald 43.
(a) What is the difference between these scores?
(b) How many more does Mike need to make 59?

Multiply or Subtract

7 In a garden there are 2 clumps each with 135 snowdrops. How many snowdrops are there?

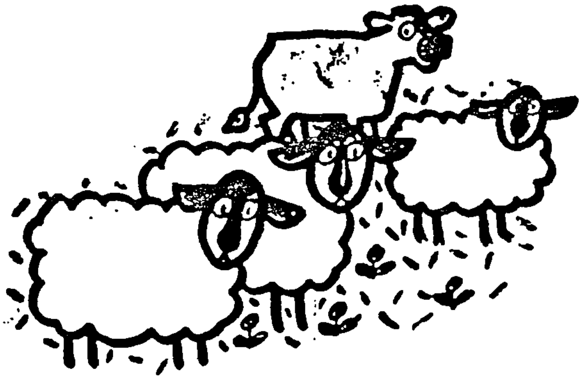
8 Jim enters the 80 metres race and is given a start of 13 metres. How far does he have to run?

Add or Subtract

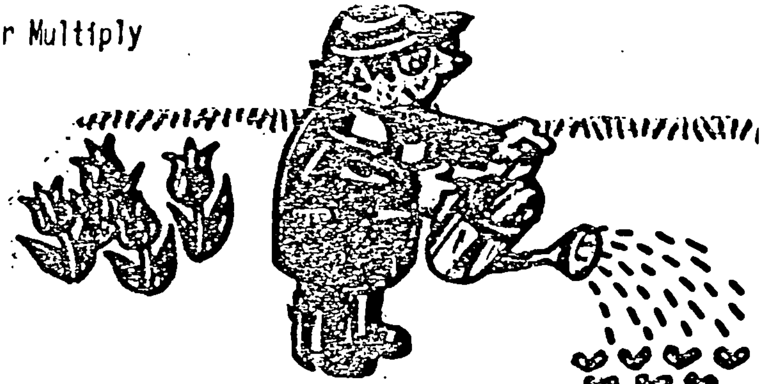
1 David enters the 85 metres race and is given a start of 13 metres. How far does he have to run?

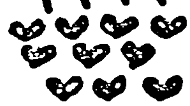
2 $\begin{array}{r} 114 \\ +\ 33 \\ \hline \end{array}$

3 Farmer Macdonald had 60 sheep. At the market he sold 5 and bought 3. How many sheep has he now?



Subtract or Multiply

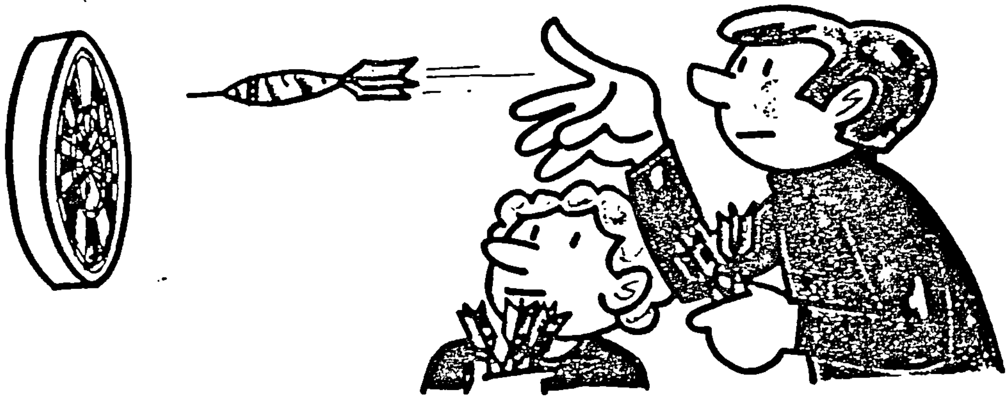




In a garden there are 2 clumps each with 125 snowdrops, 4 beds each with 69 pansies, and 6 beds each with 73 roses.

4 (a) How many snowdrops are there?
  (b) How many pansies are there?
  (c) How many roses are there?

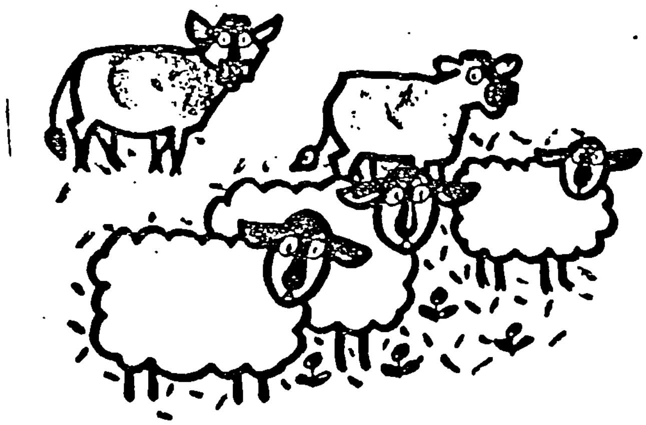
5 $\begin{array}{r} 402 \\ -237 \\ \hline \end{array}$



6 During a game of darts Chris Smith had scored 123 and Bill Brown 235.
  (a) What is the difference between these scores?
  (b) How many more does Chris need to make 255?

Add or Subtract

7 Farmer Till had 210 sheep. At the market he sold 88 and bought 25.
How many sheep has he now?



One morning 37 boys and 46 girls go to the library.
That afternoon 39 boys and 59 girls go.

8 How many (a) boys, (b) girls go to the library that day?